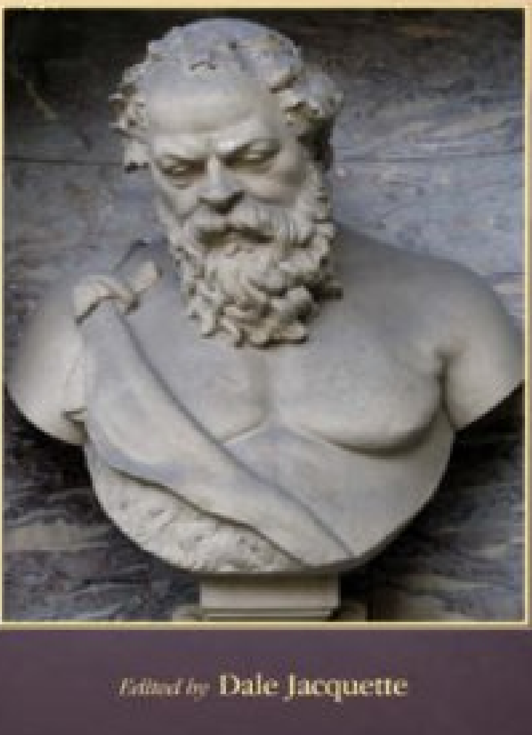
Handbook of
THE PHILOSOPHY OF SCIENCE

General Editors: DOV M. GABBAY, PAUL THAGARD, AND JOHN WOODS

PHILOSOPHY *of* LOGIC



Edited by Dale Jacquette

GENERAL PREFACE

Dov Gabbay, Paul Thagard, and John Woods

Whenever science operates at the cutting edge of what is known, it invariably runs into philosophical issues about the nature of knowledge and reality. Scientific controversies raise such questions as the relation of theory and experiment, the nature of explanation, and the extent to which science can approximate to the truth. Within particular sciences, special concerns arise about what exists and how it can be known, for example in physics about the nature of space and time, and in psychology about the nature of consciousness. Hence the philosophy of science is an essential part of the scientific investigation of the world.

In recent decades, philosophy of science has become an increasingly central part of philosophy in general. Although there are still philosophers who think that theories of knowledge and reality can be developed by pure reflection, much current philosophical work finds it necessary and valuable to take into account relevant scientific findings. For example, the philosophy of mind is now closely tied to empirical psychology, and political theory often intersects with economics. Thus philosophy of science provides a valuable bridge between philosophical and scientific inquiry.

More and more, the philosophy of science concerns itself not just with general issues about the nature and validity of science, but especially with particular issues that arise in specific sciences. Accordingly, we have organized this Handbook into many volumes reflecting the full range of current research in the philosophy of science. We invited volume editors who are fully involved in the specific sciences, and are delighted that they have solicited contributions by scientifically-informed philosophers and (in a few cases) philosophically-informed scientists. The result is the most comprehensive review ever provided of the philosophy of science.

Here are the volumes in the Handbook:

Philosophy of Science: Focal Issues, edited by Theo Kuipers.

Philosophy of Physics, edited by Jeremy Butterfield and John Earman.

Philosophy of Biology, edited by Mohan Matthen and Christopher Stephens.

Philosophy of Mathematics, edited by Andrew Irvine.

Philosophy of Logic, edited by Dale Jacquette.

Philosophy of Chemistry and Pharmacology, edited by Andrea Woody and Robin Hendry.

Philosophy of Statistics, edited by Prasanta S. Bandyopadhyay and Malcolm Forster.

Philosophy of Information, edited by Pieter Adriaans and Johan van Benthem.

Philosophy of Technological Sciences, edited by Anthonie Meijers.

Philosophy of Complex Systems, edited by Cliff Hooker and John Collier.

Philosophy of Earth Systems Science, edited by Bryson Brown and Kent Peacock.

Philosophy of Psychology and Cognitive Science, edited by Paul Thagard.

Philosophy of Economics, edited by Uskali Mäki.

Philosophy of Linguistics, edited by Martin Stokhof and Jeroen Groenendijk.

Philosophy of Anthropology and Sociology, edited by Stephen Turner and Mark Risjord.

Philosophy of Medicine, edited by Fred Gifford.

Details about the contents and publishing schedule of the volumes can be found at http://www.johnwoods.ca/HPS/.

As general editors, we are extremely grateful to the volume editors for arranging such a distinguished array of contributors and for managing their contributions. Production of these volumes has been a huge enterprise, and our warmest thanks go to Jane Spurr and Carol Woods for putting them together. Thanks also to Andy Deelen and Arjen Sevenster at Elsevier for their support and direction.

PREFACE

It was a great pleasure to be invited by Dov Gabbay, Paul Thagard and John Woods to edit this volume of *Philosophy of Logic* for the *Handbook of the Philosophy of Science* at North-Holland Elsevier. The contributors to this collection of newly commissioned essays on contemporary philosophy of logic have done superb work in covering a wide range of topics in mathematical and philosophical logic in great depth and with exemplary technical expertise.

The papers taken as a whole represent state of the art research in contemporary logic and philosophy of logic. The book serves two purposes. It provides essential background to major developments in philosophical logic for students and professional researchers, and it brings together new perspectives, new challenges and recent developments in logic and its philosophy as a sampling of current logical practice and a weathervane of logic's bright future. While no selection of writings that will fit between two covers on a library shelf can possibly hope to do full justice to a field of investigation as rich, far-reaching, and constantly changing as philosophy of logic, the present edition of papers with its particular choice of topics and illustrations of problems and methods typifies some of the most exciting work that is currently being done in some of the most important subdisciplines of formal symbolic logic. The philosophical issues surrounding new theoretical discoveries and applications of logic have implications that go far beyond mathematics. They touch base with fundamental concepts of metaphysics, theory of knowledge, philosophy of science, philosophy of language, and the values, especially aesthetic norms, that are reflected in the ways logical systems are constructed and evaluated.

An editor can only create a space in which a talented group of authors can present their ideas. I am delighted now to share the results with a readership that I hope will find their discussions as insightful as I have. If the papers collected here are deemed useful to others interested in the interplay of contemporary logic and philosophy, then the book will have achieved its purpose. All credit then is due individually to the logicians and philosophers with whom it has been my privilege to collaborate. I am grateful, therefore, in the first instance, to the authors for their excellent work. Special thanks are due to Jane Spurr, at the Department of Computer Science, King's College, London, for her professional assistance in bringing this project to publication. I am indebted to John and Carol Woods for timely and meaningful help with logistics difficulties in the early production phases of the book when I was away from my office overseas. Finally, I wish to thank the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS), Royal Netherlands Academy of Arts and Sciences (KNAW), for

a Research Fellowship in Residence in 2005–2006 that funded this among related research projects in the philosophy of logic and mathematics. I dedicate this book with love to my wife Tina and our new grandbaby David Owen Templeton.

Dale Jacquette
The Pennsylvania State University
Netherlands Institute for Advanced Study (NIAS)

CONTRIBUTORS

Timothy Bays
Department of Philosophy, University of Notre Dame, Notre Dame, IN 46556, USA.
bays.5@nd.edu

JC Beall
Department of Philosophy, University of Connecticut, Storrs, CT 06269-2054, USA.
jc.beall@uconn.edu

Katalin Bimbó
School of Informatics, Indiana University, Bloomington, IN 47408 , USA.
kbimbo@indiana.edu

Otávio Bueno
Department of Philosophy, University of Miami, Coral Gables, FL 33124-4670, USA.
otaviobueno@mac.com

Stewart Candlish
M207: Philosophy, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia.
candlish@arts.uwa.edu.au

Charles S. Chihara
Department of Philosophy, University of California, Berkeley, CA 94720-2390, USA.
charles1@socrates.berkeley.edu

John Collins
Department of Philosophy, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027, USA.
jdc9@columbia.edu

Nic Damnjanovic
The University of Colorado, Boulder, Colorado 80309-0232, USA.
damnjano@colorado.edu

John W. Dawson, Jr.
Penn State University, 1031 Edgecomb Avenue, York, PA 17403, USA.
jwd7@psu.edu

Newton CA da Costa
Department of Philosophy, Federal University of Santa Catarina, Florianópolis, SC 88040-900, Brazil.
ncacosta@usp.br

Heinz-Dieter Ebbinghaus
Albert-Ludwigs-Universität Freiburg, Mathematisches Institut, Abteilung für mathematische Logik, Eckerstraße 1, 79104 Freiburg, Germany.
hde@math.uni-freiburg.de

Peter Fletcher
School of Computing and Mathematics, Keele University, Keele, Staffordshire, ST5 5BG, UK.
p.fletcher@maths.keele.ac.uk

Siegfried Gottwald
Institut für Logik und Wissenschaftstheorie, Universität Leipzig, Beethovenstraße 15, 04107 Leipzig, Germany.
gottwald@uni-leipzig.de

Michael Hand
Department of Philosophy, Texas A&M University, College Station, TX 77843-4237, USA.
mhand@tamu.edu

Jaakko Hintikka
Boston University Department of Philosophy, 745 Commonwealth Avenue, Boston, Massachusetts 02215, USA.
hintikka@bu.edu

David Hitchcock
Department of Philosophy, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada.
hitchckd@mcmaster.ca

Herbert Hochberg
Department of Philosophy, The University of Texas at Austin, 1 University Station, Austin, TX 78712, USA.
hochberg@mail.utexas.edu

Wilfrid Hodges
School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK.
w.hodges@qmul.ac.uk

Dale Jacquette
Department of Philosophy, The Pennsylvania State University, 246 Sparks Building, University Park, PA 16802, USA.
dlj4@psu.edu

Charles F. Kielkopf
Department of Philosophy, The Ohio State University, 350 University Hall, 230 North Oval Mall, Columbus, OH 43210-1365, USA.
kielkopf.1@osu.edu

Marcus Kracht
Department of Linguistics, UCLA, 405 Hilgard Avenue, 3125 Campbell Hall, Box 951543, Los Angeles, CA 90095-1543, USA.
kracht@humnet.ucla.edu

Décio Krause
Department of Philosophy, Federal University of Santa Catarina, Florianópolis, SC 88040-900, Brazil.
dkrause@click21.com.br

Oliver Kutz
School of Computer Science, The University of Manchester, Kilburn Building, Oxford Road, Manchester M13 9PL, UK.
kutz@cs.man.ac.uk

Ernest Lepore
Rutgers Center for Cognitive Science (RuCCS), Rutgers University - New Brunswick, 152 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA.
lepore@ruccs.rutgers.edu

John Nolt
Department of Philosophy, 801 McClung Tower, University of Tennessee, Knoxville, Tennessee, 37996-0480, USA.
nolt@utk.edu

Jaroslav Peregrin
Filozofická fakulta nám, Jana Palacha 2, Praha 1, 116 38 Czech Republic.
jaroslav.peregrin@ff.cuni.cz

Nicholas Rescher
Department of Philosophy, University of Pittsburgh, 1012 Cathedral of Learning, Pittsburgh PA 15260, USA.
rescher@pitt.edu

Gabriel Sandu
Department of Philosophy, 00014 University of Helsinki, Finland.
sandu@mappi.helsinki.fi

Roy A. Sorensen
Department of Philosophy, Dartmouth College, Hanover, NH 03755, USA.
roy.sorensen@dartmouth.edu

Matthew Stone
Rutgers Center for Cognitive Science (RuCCS), Rutgers University - New Brunswick, 152 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA.
matthew.stone@rutgers.edu

Neil Tennant
Department of Philosophy, Ohio State University, 350 University Hall, 230 North Oval Mall, Columbus, OH 43210, USA.
tennant.9@osu.edu

Johan van Benthem
Institute for Logic, Language & Computation, (ILLC), University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, Netherlands.
johan@science.uva.nl
and Department of Philosophy, Stanford University, Stanford, CA 94305, USA.
johan@csli.stanford.edu

John Woods
Department of Philosophy, University of British Columbia, 1866 Main Mall E370, Vancouver BC Canada V6T 1Z1.
jhwoods@interchange.ubc.ca.

Richard Zach
Department of Philosophy, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.
rzach@ucalgary.ca

INTRODUCTION: PHILOSOPHY OF LOGIC TODAY

Dale Jacquette

LOGIC AND PHILOSOPHY

It is standard in works dedicated to topics in philosophy and logic to distinguish between philosophy of logic and philosophical logic. There is no universally agreed upon way of dividing up these categories, and some authors prefer to conflate the two and treat the two phrases as more or less synonymous.

I do not have particularly strong reasons for distinguishing or running these concepts together. If pressed, I favor a terminology in which philosophical logic involves applications of any recognized methods of logic to philosophical problems or for purposes of advancing philosophical discussions, whereas philosophy of logic is the consideration of philosophical problems that arise in relation to work in any recognized branch of logic. By this stipulation, all of the papers in this volume belong squarely in the category of philosophical logic, and thereby also fall under the category of philosophy of logic so defined. The essays all raise problems for philosophical reflection at the same time that they illustrate the methods of logic in particular applications. By addressing the nature and limits of logic from both a theoretical and practical point of view, in the context of explaining the wide variety of different types of logic that have emerged in the recent proliferation of formal logical systems, the papers directly and indirectly engage central issues in the philosophy of logic while commenting on the principles of logic and its uses primarily in philosophy, mathematics and science.

Logic, like any other subject, depends on philosophical presuppositions. It raises conceptual and especially foundational questions that can only be profitably engaged by thinkers with rigorous philosophical training. The fact that so many logicians have also been philosophers, and that the motivation for so many important innovations in logic have derived from philosophical concerns about the clarification of concepts, problems and distinctions, means that logic and philosophy of logic are historically at least if not also thematically inseparably intertwined. It is only by understanding the complex justifications for formal logical apparatus, for the exact expression of ideas and deductively valid inference of sentences, that we can hope to fathom the explosion of logical systems that have appeared since the late nineteenth century through the latter half of the twentieth and continuing now into the twenty-first century. With every new addition to the vocabulary and

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

mechanisms of formal symbolic logic there arise an exponentially greater number of philosophical questions about the meaning and implications of what logic introduces to the realms of mathematics and science.

IN THE LOGIC CANDY STORE

I want to introduce this handbook of essays in the philosophy of logic by raising a problem that I think holds a vital key to understanding the nature of logic in contemporary philosophy and mathematics. Not all of the papers in this volume do not directly address the question, but they provide ample raw material for reflecting on the kinds of answers that might be given.

The puzzle in a nutshell is to understand how it is that so many different kinds of formal systems can all deserve to be called logics. Another way to put the problem is to ask how in the first place there can possibly be a multiplicity of logical systems as opposed to just one universal logic. If we can explain what makes all of these systems logics, why each is a logic, we should have a good handle on the question of what logic itself is, what constitutes a logic and what is meant by the concept of logic.

For many centuries, in Western thought, there was essentially only one logic. One system of logic only was recognized as properly so-called. Kant, in the *Critique of Pure Reason*, Preface to the Second Edition, famously remarks: ‘That logic has already, from the earliest times, proceeded upon this sure path is evidenced by the fact that since Aristotle it has not required to retrace a single step, unless, indeed, we care to count as improvements the removal of certain needless subtleties or the clearer exposition of its recognized teaching, features which concern the elegance rather than the certainty of the science. It is remarkable also that to the present day this logic has not been able to advance a single step, and is thus to all appearance a closed and completed body of doctrine’ (Bviii). Kant in the mid-eighteenth century, after more than two thousand years after Aristotle’s time, could still truthfully maintain that there was only one logic —Aristotelian syllogistic. Yet it took only approximately another one hundred years for logicians, including George Boole, William Hamilton, Bernard Bolzano and Franz Brentano to recommend significant corrections, amendments and improvements to syllogistic logic. The beginnings were modest. First on the scene was a recognition of the need for arguments to contain four terms rather than only three. The use of syllogistic logic was always a Procrustean bed, stretching, or, more often, lopping off parts of arguments expressed informally in natural language in order to fit the regimented structure of the neo-classical syllogism. Accuracy and generality was always a problem in Aristotelian logic, and the need for a more flexible and adaptable logic, one that had no specific limitations as to number of terms, number of assumptions or conclusions, with provisions for differences in the scope of logical connectives, started to make itself felt at many levels once the grip of syllogisms on the logical and philosophical imagination started to decline, and another way of thinking about the logic of deductive inference eventually took its place.

When we stop to consider its history, today, with our understanding of logic as it has evolved, it seems strange that Aristotelian syllogistic logic should ever have been developed in the first place. It does not appear natural or inevitable in anything like the way that basic arithmetic or elementary geometry do. Despite the evolution of notations and methods, these staples of formal reasoning appear obvious and intuitively suited as formal methods of expression and calculation for their respective fields of application. They have also stood the test of time much better than syllogistic logic, even if they are no longer regarded with the same apodictic respect they once enjoyed, and even if they have been supplemented by new mathematical techniques, new algorithms and alternative axiom schemata. In the case of syllogistic logic, there are certainly inferences that conform exactly to a three-term three-sentence deduction in a particular mood and figure, with a major and minor premise, middle term and the like, exactly as prescribed by Aristotle's *Prior Analytics.* What, however, could have compelled Aristotle, not generally formalist in his thinking, to suppose that all arguments had to adhere to just the 15 valid forms (assuming the denial of existential import) of the total 256 forms combinatorially available to the syllogism under its classical restrictions? Why would anyone have agreed that all of reasoning could be reconstructed within the limited forms of the syllogism?

We need not answer these questions here. The point is only to understand syllogistic logic as firmly entrenched in the scientific-philosophical curriculum, with no competitors in the marketplace of ideas to dispute its monopoly for more than a thousand years. With this picture in mind, we turn next to the developments of the late nineteenth century, beginning with the publication in 1874 of Boole's *The Mathematical Analysis of Logic*, and in 1879 of Gottlob Frege's *Begriffsschrift* as major watersheds in the history of logic. In these seminal works, we witness the replacement of Aristotle's rigid syllogistic logic with a more expressively flexible and deductively transparent graphic language for exhibiting logical structures. It was a long time coming and a welcome new development on the symbolic logic landscape. It is surely no accident, in view of the comparison between classical mathematics and syllogistic logic that Boole and Frege should have independently found inspiration for their revisionary formulations of logic respectively in algebra and arithmetic. It still does not account for the cacophony of different and distinct logics that seem to have replaced the search for the one and only logic as the unified single formal theory of deductive thinking. For, as we move from Frege's *Begriffsschrift* to Whitehead and Russell's *Principia Mathematica* and beyond to the situation in logic today, we find ourselves with an embarrassment of logical riches.

How we got from Frege to the current state of things is an interesting historical question. One possibility is the following. I think that the most likely explanation must weave together at least three factors. One is the parallel discovery of non-Euclidean geometries by Nikolai Ivanovich Lobachevsky (1829) and Georg Friedrich Bernhard Riemann (1854) that opened the door to the possibility of alternative axiom systems for the same apparently objective body of relations in mathematics. The existence, first, and, finally, the dramatically successful application of Reimannian non-Euclidean geometry in Einstein's theory of relativity, freed the imagination of many mathematicians and logicians, who began rethinking the concept of a formal system as any self-consistent axiom system whose implications can be rigorously investigated. This paved the way for logicians to reflect in the same expansive terms about the possibilities of experimenting with different logic systems and thinking of non-Aristotelian and eventually non-Fregean and non-Russellian logics. By itself, however, the emergence of non-Euclidean geometries does not seem adequate to explain the development of alternative logics. After all, it would have been possible philosophically for many logicians to have drawn a sharp distinction between logic and mathematics, or between logic and arithmetic on the one hand, and geometries on the other, such that logic or logic and arithmetic unlike geometry are not subject to deviant axiomatizations.

What seems to have added further impetus to the onset of so many different kinds of nonstandard nonclassical logics are several philosophical motivations. These include the study of modal logics in the first third of the twentieth century, gaining momentum especially beginning in the late 1950s, and of relevance logics to overcome the paradoxes of material and strict implication in the 1960s and 1970s, all to the present day. The formalization of model set theoretical semantics for modal logics in particular opened the door to today's expansive proliferation of logical systems. It did so, moreover, in several specific ways. First, the formal semantics developed for multiple modal systems was already in itself a pluralization of logical axiomatizations. The differences in theorems and valid inference structures for different permissible iterations and inferences from iterated quantified modal logics immediately swelled the ranks of logic with a host of new formal systems. Second, the existence and exact interpretation of distinct modal logics encouraged logicians and philosophers to expand their horizons with respect to other areas of discourse that had resisted analysis in classical logic. This included most notably the development of relevance logics. From an altogether independent direction and with its own particular motivations, efforts to better understand entailment relations in such a way as to avoid trivial implications involving contingently or necessarily false assumptions and true or necessarily true conclusions, in effect, investigating the logic of relevant implication, resulted in the formalization of another burgeoning family of logics. The same can be said for intuitionistic logics, beginning with L.E.J. Brouwer's work in 1907 and continuing especially from the 1930s to today.

The significance of these early defining moments in the recent history of symbolic logic is that philosophical considerations have driven a desire to extend the palette of formal logical systems. These have taken root because of a perceived need to make logic more expressively adapted to specific areas of linguistic usage inadequately served by existing classical logics, or in order to overcome formal or epistemic scruples about the proper modeling of deductive inference. Innovations from a related direction have also come from the desire to determine exactly the scope and limits of logical symbolisms in formal metatheory and metamathematics. A notable example, of which many might be given, is the discovery of the arithmetization of logical syntax in Kurt Gödel's 1931 proof of the deductive incompleteness of classical first-order logic. All such trends, it should be needless to say, are not only justified but commendable and even indispensable to the health and vitality of mathematical logic as a living evolving tradition in the formal sciences.

UNDERSTANDING THE PLURALITY OF LOGICS

The more recent extensive growth of symbolic logics makes it harder than ever to sustain the attitude that logicians are in search of a single formalization that will stand as the logic of all thought and discourse. What is suggested in the spirit of modern specialization is a plethora of different logics, many of which are deductively incompatible with one another, some of which are at most overlapping in their theorems and algorithmic methods or in some of their results, and some that are so distant from one another that it is hard to see the sense in which they can all claim to be logics.

This, of course, returns us to the problem at hand. How is it that all of these formal systems can rightly claim to be logics? What can we learn about the concept of logic itself that for the first time in its history is so diverse in its formulations across the board in mathematics and philosophy? There are several possibilities that immediately come to mind, and it may be worthwhile to display them systematically for the sake of comparison.

1. There is a universal logic. The alternative formalisms that are currently on offer as logics are in fact something other than logic. These formal languages may eventually find a distinct name and separate category. The proliferation of what are presented as different logics can accordingly be understood as a sometimes deliberate and sometimes subconscious attempt on the part of logicians to identify the one and only correct logic that has yet to emerge as the legitimate successor of syllogistic logic.

2. All of the various systems of logic that seem so disjoint today will eventually or can potentially be unified and integrated as a single super-system. Some systems that do not lend themselves to such incorporation may fall by the wayside as faux logics. The contemporary proliferation of distinct logics is comparable to the prior systematization of other mathematical disciplines like Euclid's geometry, that developed out of a hodgepodge of practical methods of measurement and calculation recipes that were finally brought together under a single set of axioms. We are, speaking historically, currently in a pre-systematization phase in the development of symbolic logic, working piecemeal on the fragments of what will eventually emerge as a grand synthesis in a future logic of enormous power and comprehension.

3. There is no universal logic, but there is a universal concept of logic that

underlies one and all of the many different logical formalisms that have been advanced. On this view, it becomes the task of the philosophy of logic to analyze and define the concept of logic in terms of its necessary and sufficient conditions in such a way that all of the logics properly so-called will properly fit within its conditions.

4. There is a universal logic and a universal concept of logic, but we can never fully or confidently come to know what it is. We are thus in a situation that is similar to that of Socrates in Plato's early aporetic dialogues, in which a concept persistently eludes us as we try to nail down its necessary and sufficient conditions. What we do instead, as the next best choice, is to propound many different formal systems, each of which purports to be a logic, and some of which may purport to be *the* logic, but without any conceptual guidance as to how closely or distantly any chosen formalism might be approximating the ideal. The proliferation of 'logics' is thus itself a symptom of our lack of a conceptual polestar. The best approximation might then be regarded as the one, if any, that gains the most widespread acceptance or turns out to have the most fruitful applications.

5. There is no universal logic or universal concept of logic, but something rather like a later Wittgensteinian family resemblance among disparate logics. The overlappings of one logical system with respect to another in the entire network of logics must then be carefully mapped out so that we can understand the conceptual interrelations between different kinds of logic, in addition to grasping their historical-genealogical and thematic interconnections. Logics, as Wittgenstein said of mathematics, and no doubt also with symbolic logics in mind, are in that case rather like a city with a central core and suburbs burgeoning in various directions to serve a variety of different kinds of expressive and inferential needs, with no prospect of unified integration into a single super-system.

6. There is no universal logic, universal or family resemblance concept of logic that can hope to incorporate all of the disparate systems of logic. These systems are rightly called logics only nominalistically, in that they are formal systems developed and referred to institutionally as logics by professional logicians. There is no possibility of such logics being integrated, and no prospect of understanding any deeper connections by virtue of which they are related. They are all logics simply because they are given that name by persons who are part of a professional scientific and philosophical community with the institutional authority to designate them as such. In the past, there was only one logic because logicians were too limited in their imaginations or too hidebound by tradition to venture out in different directions. The idea of a unified integrated logic or overlapping family resemblances among logics is a myth lacking any practical foundation.

We need not assume that these are the only possibilities. They nevertheless represent a reasonable place to begin in thinking about the problems posed for philosophy of logic, given the flourishing today of so many apparently different logics. If we try to take stock of some of the most conspicuous differences that divide these logics, we might be struck at once by the fact that new logics have not only sought to fill in some of the gaps left by the early pioneering formalisms of Frege and Whitehead and Russell, but have developed as a response to the desire to better understand the nature of deductive inference.

Thus, it is one thing to supplement a generically *Principia*-style mathematical logic with provisions for modal operators, alethic, deontic, temporal, and others, or with devices for formalizing indexicals and dealing with anaphoric reference in colloquial language, to provide for a logic of imperatives or interrogatives modeled on that of the logic of declarative propositions, or the like. Other logical systems, however, have expressly been designed with the intention of addressing philosophical concerns about the logic of deductive inference in its purest most general terms. Here we may naturally think of intuitionistic logics, relevance and paraconsistent logics, along with a spectrum of nonstandard philosophically motivated semantics for standard predicate-quantificational logics, including substitutional semantics, probability and game theoretical semantics, intensional systems, free logics, logics of nonexistent objects, logics of beingless objects, independence friendly logics, and many others besides, with no obvious limit to their proliferation.

What is fascinating about these logics is the ways in which they challenge our understanding of what is and is not required for deductive validity. It is one thing to claim that a deductively valid inference is one in which if the assumptions are true then the conclusions must be true, or such that it is impossible for the assumptions to be true and the conclusions false. It is quite another matter to specify how we are to handle situations in which the assumptions cannot be true or the conclusions cannot be false, even if the assumptions have no shared content or point of contact with the conclusions. It is another matter yet again to decide in exactly what sense the conclusions of a deductively valid argument must be true, or such that it is impossible for them to be false, when the assumptions are true. What sort of modality is involved in deductive validity, and what is its logic? Does the modal logic in turn require a concept of deductive validity with the same or a different modality governing the semantic possibilities of modal assumptions and modal conclusions? These are among the many still unanswered questions in philosophy of logic. Nor can we simply take it for granted that received wisdom about semantic categories of truth and falsehood carry over straightforwardly into symbolic logic. There are multiple conceptions of truth and falsehood, with distinct implications for the syntax and semantics of alternative logical symbolisms. There are systems that contrive to tolerate syntactical contradictions without classical inferential explosion, or that try to confine intuitively irrelevant inferences to limited parts of the range of a deductive consequence relation. There are systems that seek to avoid truth- or validity-challenging deductive paradoxes by introducing nonclassical truth values or truth value gaps, and others again that model truth value assignments on fuzzy probability coefficients. The possibilities sometimes appear to be endless, and yet all are at least popularly considered to be *logics*. Whether and in what sense they are rightly classified as such depends very much on arriving at a satisfactory answer to the question we have posed for the philosophy of logic.

Nevertheless, there are constraints operating within each logical system. Efforts to construct logical formalisms run into limitations, regardless of their starting places; there are always things they can reasonably be made to do, and things they cannot do. It is tempting as a result to suppose that logic after all has something to do universally with the limitations encountered in different ways and in different places by all of these symbolisms. The study of alternative logical systems could then be instrumental after all in touching base with a universal concept of logic. We can begin in this way to grasp a sense of what is meant by logic, even if no single formalization of logical reasoning can be expected to explicate what is common to all deductively valid inference. It might nevertheless be objected that if in fact there are different concepts of deduction, then there must be different concepts of deductive logic, and hence different concepts of logic. Does this mean that there cannot be a grand synthesis of logic or an underlying unified integrated universal logic? Not necessarily; it all depends on whether or not ostensibly different concepts of deductive logic can or cannot be brought together as multiple aspects of a single notion of deduction. This, of course, is an unresolved issue, but we learn more from efforts to develop the individual logics that pose the problem and from attempts, even when they spectacularly fail, to bring order and system to the diversity of formalisms that continue to flood the market. Whether or not such projects in the philosophy of logic have any chance of succeeding is currently unclear. It is easy to imagine many logicians and philosophers not being particularly interested in the problem, but keeping their noses to the grindstone and working exclusively on the individual challenges posed by hammering out the details of particular logics dedicated to particular analytic tasks.

The variety of logical systems under development contribute in different ways to understanding the nature of logic, even when they are not undertaken with any philosophical agenda or interest in the philosophy of logic on the part of their authors. They provide the unselfconscious raw material of philosophical logic, for explicating the concept of deductively valid inference in previously undreamed of ways. In the field of symbolic logic today, with so many different formal systems from which to choose, we can revel like children turned loose in a shop selecting their favorite sweets. If that is the new disjoint, decentralized model for research in logic, if we do not have to worry about whether or how a preferred logic relates to *logic* or to *the* one and only true system of logic, then the philosophy of logic needs to take account of these developments and try to learn what it can about the concept of logic from the surfeit of logical systems originating during this time of extensive exploration and experimentation.

FUTURE DIRECTIONS IN PHILOSOPHICAL LOGIC AND PHILOSOPHY OF LOGIC

The present volume offers a snapshot of contemporary developments in philosophical logic. The philosophy of logic, like the philosophy of art or philosophy of science, takes its bearings from the best work in the field.

The handbook includes exemplars of new findings in logical theory and its applications, and provides a valuable compendium of resource materials. It includes studies relating to the background of informal logic, semantic analysis, logical paradoxes, formal explications of the concept of truth, deductive validity and logical entailment, deductivism and Hilbert's program in mathematical logic, the theorems of classical metatheory, cardinalities of semantic domains for the objectual interpretation of quantifiers, objectual versus substitutional interpretations of the quantifiers in predicate logic, many-valued logics, relevance logics, paraconsistent logics, free logics, extensional versus intensional logics, the logic of fiction and reference in logic to nonexistent and beingless objects, analysis of vagueness and the relation of symbolic logic to ordinary colloquial language, models of modality, logically possible worlds, transworld identity and counterpart semantics for modal logic, the controversy surrounding the ontology of logically possible worlds, modal realism versus modal actualism and modal fictionalism, the logic of counterfactuals and analysis of the concept of causation and causal relations, the history and continuing prospects of logicism in the reduction of all of mathematics to logic, the concept of infinity and transfinite domains in logic and mathematics, the practical uses of logic in mathematics, epistemology, and the natural sciences, and general philosophical reflections on the nature of logic, its scope and limits and philosophical applications.

A handbook is meant to be of such size that one can take it up and hold it in the hand, to flip through its pages and browse its chapters for subjects of interest and refer to it for essential information. Within these limitations, the guiding idea of the present handbook has been to provide an overview of contemporary research in logic for future directions in its philosophy. Where will philosophy of logic go next? We have presented some possible lines of inquiry, suggested by the puzzle of trying to understand the nature of logic in light of the abundance of logical systems. The lack of universal agreement about even the most basic logical concepts, reflected in alternative logics of deductive inference and incompatible interpretations of fundamental semantic concepts around which different logical formalisms have been devised, pose an interesting analytic challenge for the philosophy of logic. We can imagine that some of its future directions might stem ultimately from a desire to clarify the concept of logic itself and to come to terms with the vast plurality of formal logical systems. We might expect that some thinkers will try to eliminate what they deem as excesses among logical formalisms, leading toward a new austerity in philosophical logic, trimming down its ambitions in response to a more circumscribed conception of what is to count as logic, logical deduction, and the proper ambit of deductive reasoning. Others will further elaborate platonistic, conceptual, constructivistic, intuitionistic, or similar traditional ontologies for philosophical logic, with implications for what is to count as a genuine system of logic and how interrelations among logical formalisms should be understood. Still others will develop latitudinarian Wittgensteinian or nominalistic philosophies of logic that embrace all or virtually all work ostensibly in the area of logic as exemplifying genuine logical systems, and testifying to the impossibility of identifying any meaningfully universal characteristics of logic.

All this, it will be said, is already known. The question is, Are there discernible trends that are beginning to emerge in current logical theory that the essays collected in the present volume help to exemplify or to which they contribute? Here I think that the only responsible thing to say is that the jury is still out and that no clear indication of a verdict is yet to be divined. If the essays collected in the present volume encourage any predictions about the future course in the philosophy of logic, the following might be the most that can responsibly be projected, subject to revision. Reading the tea leaves, one possibility strongly suggests itself. This is that logic will continue to diversify itself in many different directions, that there will be a flowering of new systems of logic formulated in order to explore and test the viability of philosophical concepts of deductive inference and the semantics of thought and language, while continuing to address areas of discourse considered to be inadequately served by logical systems currently available. There will be no dramatic integration or unification of these disparate logics for the foreseeable future, although linkages and conceptual connections, shared formalisms and algorithms may continue to suggest the possibility of drawing all of these logics together into one comprehensive system. At best, we might look forward to a process of coalescence and affinity recognition that will only occur in limited ways for specific parts of logic. Constellations of logical systems will then begin to take shape and separate themselves into distinct clusters of related logics. The semantics, analysis of the concepts of truth and falsehood, and accompanying metaphysics for such diverse logics will begin to diverge from the purely extensionalist semantics that have held sway since the time of Bertrand Russell and W.V.O. Quine, and start increasingly, as with Donald Davidson, to return to a more Fregean combined intensional and extensional theory, favoring interpretations based on the possibilities of combinations of properties to constitute the sense of referring terms and sentences, and to stand in relation to the referents to object terms to which properties are predicated in true and false sentences.

If a sweeping integration of the plurality of logics is to be effected, it will most likely occur in part at least, like the solution to the four-color problem in topology, through the mechanization of logical syntax and inference rules in automated reasoning systems. Computers are capable of comparing astronomical numbers of theorems of astronomical numbers of different logics with heuristics for identifying commonalities. As a tool for investigating the diversity of logics, computerized comparisons and taxonomies of formal languages may increasingly reveal underlying structures or the lack thereof that in turn can help to answer fundamental problems in the philosophy of logic. The mere fact that the syntax of logical systems can be automated by itself in the meantime will not answer deeper questions in the philosophy of logic concerning the semantics and ontology of logic, nor will it help us better to understand the nature of logic itself, whether platonistic, nominalistic, intutionistic, pragmatic, or something altogether different, previously unknown on the horizon.

Although computers cannot answer these questions for us, they can nevertheless provide valuable instruments to aid in the task. We can use information processing machines to assemble and analyze data in order to focus attention on aspects of longstanding philosophical questions that would never be possible relying entirely on human brain power. We need all the mechanical help we can get in assessing the interrelations among formal logical systems from a variety of standpoints, in order to uncover and investigate properties of logics that would otherwise remain invisible to the plodding efforts, limited memories and cognitive workspace of human inquirers. Ongoing efforts to automate reasoning are likely, therefore, to play an increasingly important role in this type of research. Such projects are worthwhile for their own sake in numerous applications of symbolic logic in technology and the information sciences. They provide a powerful adjunct in the effort to gain perspective on outstanding problems in the philosophy of logic — problems that in some ways have been with us since the time of the ancient Greeks, and that in other ways arise uniquely only today in the contemporary logic candy store. It is clear that in this work, if we are not to fall victim to a kind of vicious circularity, we will need to pay careful attention to the kinds of logic embodied in the hardware and software built into automated systems, and the presuppositions and limitations of those logics, insofar as they are used to evaluate the interrelations among symbolic logical systems generally. Could this then be the path toward a universal logic, developing a single computer logic for the purpose of electronically comparing intersystemic logical formalisms? Will computers in turn be needed to devise an automated logical system for purposes of comparing, evaluating, and seeking out similarities among other logical systems? What we will potentially gain from the future automation of the wide diversity of logics, developed by hand in the time-approved manner that began with Aristotle, is a unique window on the nature of logic, if, indeed, logic has an underlying nature, and on the concept of logic itself, or, alternatively, on the hard-won insight that there is after all no universal nature of logic, nothing whatsoever in common between the variety of formal systems that are authoritatively designated as systems of formal symbolic logic, beyond the fact that they are so designated.

It is with an eye toward these possible future developments in the philosophy of logic that the present volume has in part been organized. The handbook, if it is to be useful not only immediately, but for generations of logicians and philosophers to come, must reflect a vision of the current state of research in philosophical and mathematical logic that is compatible with an open-ended sense of how things are likely to proceed in the indefinite future. Like most interesting questions in philosophy, the topics investigated in the philosophy of logic, as logic and philosophy continue to mature together in symbiotic interrelation, are certain to remain hotly disputed, and, to many minds, indefinitely unresolved. The longevity of such philosophical challenges does not in the least detract from the importance of the topics explored, but adds an element of open-endedness to the inquiry that continues to contribute to the production of new insights, new meanings, and even to unanticipated distinctions and technical innovations. That itinerary, in any case, is a (virtually) guaranteed implication for philosophy of logic in the future.

WHAT IS LOGIC?

Jaakko Hintikka and Gabriel Sandu

1 LOGIC AND INFERENCE

It is far from clear what is meant by logic or what should be meant by it. It is nevertheless reasonable to identify logic as the study of inferences and inferential relations. The obvious practical use of logic is in any case to help us to reason well, to draw good inferences. And the typical form the theory of any part of logic seems to be a set of rules of inference.

This answer already introduces some structure into a discussion of the nature of logic, for in an inference we can distinguish the input called a premise or premises from the output known as the conclusion. The transition from a premise or a number of premises to the conclusion is governed by a rule of inference. If the inference is in accordance with the appropriate rule, it is called valid. Rules of inference are often thought of as the alpha and omega of logic. Conceiving of logic as the study of inference is nevertheless only the first approximation to the title question, in that it prompts more questions than it answers. It is not clear what counts as an inference or what a theory of such inferences might look like. What are the rules of inference based on? Where do we find them? The ultimate end one tries to reach through a series of inferences is usually supposed to be a true proposition. Frege [1970, 126] wrote that “the word ‘true’ characterizes logic.” But how does this desideratum determine the rules of inference? A few distinctions will illustrate the embarrassment of riches covered by the term “logic” and at the same time sharpen the issues.

Inferences can be either deductive, that is, necessarily truth preserving, or ampliative, that is, not necessarily truth preserving. This distinction can be identified with the distinction between such steps in reasoning as do not introduce new information into one’s reasoning and such as do not do so. For if that information is genuinely new, its truth cannot be guaranteed by the old information. Or, rather, we could thus identify deductive inferences as uninformative ones if we had a viable notion of information at our disposal. Unfortunately, the viability of a notion of information that could be used to make the distinction has been challenged by philosophers, notably by W.V. Quine [1970, 3-6, 98-99], as a part of his criticism of what he calls the analytic-synthetic distinction [Quine, 1951]. A closer examination shows that we have to distinguish from each other two kinds of information, called depth information and surface information [Hintikka, 1970; forthcoming (c)]. A valid deductive inference cannot introduce new depth information, but it can

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

increase surface information. Quine's scruples can be best understood as being based on the difficulty of separating depth information and surface information purely behaviorally.

Philosophers generally consider deductive reasoning as the paradigmatic type of inference. However, there is a widespread view among nonphilosophers that might be called the Sherlock Holmes view. According to it all good reasoning including ampliative reasoning, turns on "deductions" and "logic". (Cf. [Hintikka, 2001].)

2 DEDUCTIVE INFERENCE AND INFORMATION

Being truth-preserving, deductive inference is essentially cumulative or, to use a more common term, monotonic. In recent years, spurred largely by studies of automated reasoning of different kinds, there has mushroomed a wide variety of nonmonotonic logics. (See e.g. [Antoniou, 1997; Gabbay *et l.*, 1993-1996, vol.3].) They are all obviously modes of ampliative reasoning, and should be studied as such. They will be commented on in §17 below.

If a deductive inference does not introduce new information, it is in some sense uninformative or tautological. Such uninformativeness of deductive inference was maintained by among others by the early Wittgenstein and by the logical positivists. (See [Hempel, 1945; Dreben and Floyd, 1991].) Such a view is unsatisfactory, for it leaves unexplained what the reasoner gains by drawing such allegedly uninformative conclusions. Something important can obviously be gained through deduction. This air of unintuitiveness can be dispelled by means of the distinction between depth information and surface information mentioned in §1. Surface information can be characterized as the information that can be read off from a sentence without any nontrivial deductive ado, while depth information is the totality of (surface) information that can be extracted from it by deductive logic. The emergence of new surface information explains why purely deductive inference can be experienced as informative and even productive of surprises.

The noninformativeness (in the sense of depth information) of deductive inference is a presupposition of one of the most important kinds of application of logic. This application is axiomatization. The leading idea of this method, vividly emphasized by David Hilbert, is to summarize our knowledge of some model or class of models, for instance certain physical systems, in an axiomatic theory [Hilbert, 1918]. That model or those models can then be studied by deducing theorems from the axiom system. Such a summation is not possible if new information about the subject matter can be introduced in the course of the deduction.

3 ANALYTIC INFERENCES

A distinction within deductive inferences can be made between those that depend only on the meaning of logical concepts, expressed in language by what is known as logical constants, and those that depend also on the meaning of nonlogical constants. Deductive logic in the narrow sense is restricted to the former. Truth-preserving inferences that are not purely logical in this narrow sense are often called analytic. This distinction presupposes a distinction between logical and nonlogical concepts, expressed by logical and nonlogical terms. It is not obvious how such a distinction can be drawn, except perhaps by merely enumerating the logical constants of the given language. (Cf. here [Etchemendy, 1983; 1990; Hintikka, 2004b].) (Concerning the different meanings of "analytic" in logic and its history, see [Hintikka, 1965; 1973].)

Deductive inferences involving nonlogical words are studied among other places in what is known tendentiously as philosophical logic. (This label is misleading because it suggests that deductive logic in the narrower sense is not philosophically important.) Historically studies in such earliest "philosophical" logic were logics calculated to capture such notions as necessity and possibility. They are known as modal logics. (See e.g. [von Wright, 1951a; Carnap, 1947; Kripke, 1963; Lemmon and Scott, 1977; Shapiro, 1998; Blackburn *et al.*, 2001].) Later, attempts have been made to capture such related notions as entailment and relevance. (Cf. e.g. [Anderson *et al.*, 1992].) Philosophical logic also comprises among its other branches epistemic logic, that is, the theory of inferences that depend on the meaning of epistemic terms (see [Hintikka, 1962; 2003]), deontic logic (theory of normative discourse (see [von Wright, 1951b; 1983; Hilpinen, 1971])), and tense logic (theory of inferences turning on the meaning of temporal terms (see [van Benthem, 1982; Gabbay *et al.* 1993–1996, vol. 4])).

The logical behavior of many terms that do not overtly deal with truth-preserving inferences can still be studied by the same method and often partly reduced to the theory of truth-preserving logics. Cases in point are the logic of questions and answers (erotetic logic) which can be considered an aspect of epistemic logic. This is made possible by the fact that questions can be construed as requests for information and their behavior can be studied by reference of the statement (known as the desideratum of a question) which specifies the epistemic state that the questioner wants to be brought about. Similar things can be said of the logic of commands.

Borderline cases are logics that depend on the meaning of probabilistic terms (probability logic [Adams, 1998]), or the behavior of the membership relation (set theory), or on the part-whole relation (mereology (see [Husserl, 1913; Fine, 1995; Smith, 1982])). Indeed, attempts have been made, among others by Rudolf Carnap [1950; 1952], to interpret probability (in at least one of its senses) as a purely logical concept. An especially difficult question concerns set theory. Some of the most characteristic axioms of set theory, for instance the axiom of choice, can be construed as logical truths of higher-order logic (see below). Yet in set theory models are sometimes studied in which the axiom of choice is not true. In the usual foundations of set theory the axiom of choice is not treated as a part of the logic used but as an additional mathematical assumption. Thus currently both probability theory and set theory seem to reside in a limbo between logic and special mathematical theories.

4 AMPLIATIVE INFERENCE

Ampliative reasoning includes prominently inductive reasoning, in which the premises are particular propositions and the conclusion either a generalization or a new particular proposition [Skyrms, 1966]. It is to be noted, however, that induction or *epagoge* originally included and sometimes simply meant a different kind of inference, viz. the extrapolation and interpolation of partial generalizations [Hintikka, 1993]. The logic of inductive inference in this historical sense is quite different from that of the ordinary induction which might be labelled “Humean induction”. Inductive inference in this sense can be either qualitative or probabilistic. Qualitative induction can depend on the elimination of alternatives to the conclusion (eliminative induction). Probabilistic induction is sometimes thought to subsume inductive inference to probabilistic inference.

Even though induction was earlier thought of as the only alternative to deductive reasoning, an abundance of other kinds of ampliative modes of reasoning have recently been studied. They include the theories of rational belief change and various forms of nonmonotonic reasoning. They will be discussed below in §17.

Besides the usual so-called classical logics there are also logics that do not involve nonlogical concepts but which are supposed to be grounded on a nonstandard interpretation of logical notions themselves. They include intuitionistic and constuctivistic logics. (See [Heyting, 1956; Dummett, 1977] and [Bishop, 1967].) Some logics are supposed to be applicable to special subject matters, as e.g. the so-called quantum logics. (See [Hooker, 1975–77].)

5 LOGICAL VS. EXTRALOGICAL SYSTEMATIZATION

Instead of thinking of logic as the study of inferences, it is sometimes construed as a study of logical truths. The two conceptions are closely interrelated. In order for a deductive inference from F to G to be valid, the conditional sentence If F, then G (in symbols $(F \supset G)$) must be logically true. Whether the converse relation holds depends on whether the validity of $(F \supset G)$ guarantees the existence of a rule of inference that mandates a move from F to G. Indeed, more may be required for the validity of an inference from F to G than the logical truth of $(F \supset G)$. For instance, it may be required that the rule takes us from a way of verifying F to a way of verifying G.

Speaking of logical truths can be misleading, however, for logical truth need not and perhaps should not be considered as a variant of plain ordinary truth. Truth is relative to one model (scenario, possible world), whereas logical truth is on this conception truth in all possible models (scenarios, possible worlds). This distinction is reflected in a parallel distinction between axiomatizations of ordinary substantial theories and axiomatizations of this or that part of logic. What are known as formal systems of logic are systematizations of logical truths. They are formulated in a partial analogy with substantial axiomatic theories. A few formulas or types of formulas are designated as axioms from which theorems are derived according to purely formal rules. This has been the standard format of formal systems of logic since Frege.

There are nevertheless important differences between the two kinds of axiomatizations. What an ordinary axiom system accomplishes (if successful) is to capture all the truths about the relevant system or systems, as deductive consequences of the axioms. In other words, what an ordinary non-logical axiom system is calculated to do is to help us to study some aspect of reality or some mathematical structure (or class of structures) by examining the logical consequences of an axiom system. In contrast, what a formal system of some part of logic is calculated to do is in the first place merely to list mechanically all the logical truths of that part of logic. This is a substantial difference. It is for instance not always the case that logical consequence relations can be captured by purely formal (mechanical) rules of deductive inference. Furthermore, the so-called rules of inference of a logical axiomatization need not ever be truth-preserving, unlike the real rules of inference of a substantial axiom system. They are not necessarily intended as guidelines for actually drawing inferences. In a sense, only the rules of inference of a nonlogical (substantial) axiom system are genuine rules of logical inference. Hence the role of logic as a system of rules of inference is in a sense more prominent in nonlogical axiomatizations than in logical ones.

Likewise, the exhaustiveness of the enumeration, known as completeness of the axiom system, is different in the two kinds of axiomatization. In the case of logical axiomatization, the term "semantic completeness" is typically used, whereas in the case of substantial axiom systems one usual term is "deductive completeness."

6 SYNTAX VS. SEMANTICS

The concept of logical truth becomes useful when we begin to develop an explicit logical theory, maybe even an axiomatic one. Such a systematic logical theory can be implemented in two main ways. One leading idea in the early development of modern logic was to try to invent a notation or symbolism so explicit and so flexible that the criteria of valid inference and of logical truth could be captured by reference to the formal features of this symbolism, without referring to their meaning. In other words, the application of these criteria happens by a calculation with symbols. This idea is what the term "symbolic logic" was calculated to highlight. Such a study of the purely formal features of a language is usually referred to as its syntax. In systematic logical studies this approach includes what is known as proof theory, and is sometimes referred to as such. This idea is somewhat similar to the project of transformational grammarians like Noam Chomsky to capture what are essentially semantical notions, such as ambiguity, co-reference and logical form, in purely syntactical terms [Chomsky, 1975].

In contrast, a study of the relations of a language to the reality it can be used to represent is known as its semantics. In logic, it includes what is known as model theory, even though this term usually refers to the particular kind of studies launched largely by Alfred Tarski [1935; 1944, 1956]. Sometimes these two are distinguished from pragmatics, which is supposed to be the study of the uses of language. This distinction is not well defined, however, for the language-world relations studied in semantics can be constituted by rule-governed uses of language, as for instance in Wittgenstein's language games or in game-theoretical semantics [Wittgenstein, 1953; Hintikka and Sandu, 1997].

One of the most important semantical concepts is that of truth. Since the validity of a deductive rule of inference means that is is necessarily truth preserving, the theory of deductive logic must have its foundation in a theory of semantics (model theory). At the earlier stages of the development of modern logic, the need of such a basis was usually denied or at least neglected. The hope was that in an appropriate symbolism all the important semantical properties and relations between propositions would have syntactical counterparts that could be studied proof-theoretically, for instance by means of suitable formal systems of logic.

7 THE LIMITS OF SYNTACTICAL APPROACHES

This hope was in the case of logical notions shattered by the incompleteness theorems of Kurt Gödel and the impossibility theorems of Alonzo Church and Alfred Tarski. (See [Davis, 1965].) Gödel showed that any first-order system of elementary arithmetic is inevitably incomplete in that some arithmetical propositions in it are true but unprovable. He also showed that the consistency of a system of elementary arithmetic cannot be proved in the same system. Church showed that the decision problem for first-order logic is undecidable, that is, that the set of logical truths of this part of logic is not recursive. Tarski developed techniques of defining the concept of truth for suitable logical languages, but at the same time showed that such definitions can only be formulated in a richer metalanguage [Tarski, 1935]. Since our actual working language presumably is the richest one we have at our disposal, the concept of truth cannot according to Tarski be consistently applied in what Tarski called the "colloquial language". In particular, Tarski showed that the concept of truth for any first-order axiom system cannot be defined in the same system.

These results are also relevant to the question of the relationship of syntactically explicit ("formalized") logic and the logic implicit in natural languages and in ordinary discourse. Earlier it was often thought and said that the logical symbolism is little more than a streamlined and regimented version of an ordinary language. The main advantage of a symbol language was seen in its freedom of ambiguities and other obscurities.

This idea lives on in a modified form in that some theorists of language consider the notion of logical form, expressible in the logical symbolism, as playing an important role in grammatical theory.

There is strong evidence, however, to suggest that logical symbolism and ordinary language are more radically different than was assumed by the earliest "symbolic logicians", For instance, the presuppositions of first-order language examined in SS11–12 below are fairly obviously not satisfied by our ordinary language. Likewise, the notion of independence discussed in §12 is not expressed in most natural languages in a uniform way, even though it turns out to be quite prevalent in their semantics.

The wisest course appears to be to consider formal logical languages as alternatives to one's own natural language in the same sense in which one's second language offers an alternative way of expressing oneself. What can be hoped is that there might be a privileged kind of logical language that would express the true logical forms underlying all languages, formal and natural. This ideal language would be a true "language of thought" alias Frege's *Begriffsschrift* or conceptual notation [1879]. Whether our actual logical languages can claim to be approximations to such a language of thought, and if so to what extent is a moot question.

8 THE LIMITS OF SEMANTICAL APPROACHES

The need of a semantical basis of a theory of deductive logic has nevertheless been denied for different reasons. One reason is the view which was strongly represented among the early major figures of contemporary logic (Frege, early Russell, early Wittgenstein etc.) and which was still current recently (Quine, Church etc.) to the effect that the semantics of a language cannot be expressed in the same language. This kind of view has been strongly encouraged by the result of Alfred Tarski [1935] according to which the central semantical notion of truth can be defined for a first-order language (see below) only in a richer metalanguage. In so far as this result applies to our actual working language, it suggests that the notion of truth cannot play a role in our "colloquial language" as Tarski called it, for it cannot have any richer metalanguage above it.

If this impossibility really prevails, there cannot be any general semantical foundation of logic understood as a theory of inference.

How conclusive are these negative results in a general perspective? This question will be revisited below. In any case, some parts of logic allow for a complete axiomatization. For one thing, Gödel proved the semantical completeness of the received first-order logic in 1930. Since this logic is often taken to be the central area of logic — and even taken to exhaust the purview of symbolic logic —, this has created the impression that the dream of purely symbolic deductive logic is indeed realizable and that the negative results pertain only to impure extensions of the proper province of deductive logic. In any case, the questions whether different parts of logic admit of a semantically complete axiomatization and whether different mathematical theories admit of a deductively complete axiomatization are crucial in establishing the prospects and the limits of symbolic (syntactical) methods in logic. This is why Gödel's first incompleteness theorem is so important, in that it shows that elementary arithmetic cannot be completely axiomatized in the sense of deductive completeness.

The question of axiomatization should not be confused with the question whether logic is formal discipline. By one commonly used definition, in deductive logic a sentence is logically true if and only if its truth depends only on the way in which logical terms occur in it. If so, logical truths of deductive logic are all formal independently of the question of whether they can be mechanically enumerated by means of a formal axiom system.

9 STRATEGIC VS. DEFINITORY RULES

An important further distinction between different aspects of logical studies derives from the nature of logical reasoning (inference-drawing) as a goal-directed activity. In practically all such activities a distinction can be made between different two kinds of rules. This distinction is especially clear in the case of games of strategy. (Cf. [von Neumann and Morgenstern, 1944].) In them, we can distinguish the definitory rules which specify what may happen in the game from the strategic rules which tell how to play the game better or worse. For instance, the definitory rules of chess determine what moves are possible, what counts as checking and checkmating etc., whereas following the strategic rules of chess is what makes a player better or worse. Strategic rules are not merely heuristic. They can in principle be as precise as the definitory rules, even though they are quite often so complicated as to be impossible to formulate explicitly.

Logical theory involves the study of both definitory rules and strategic rules of the "game" of logic. (See [Hintikka, 2001].) It is important to realize that what are called rules of inference must be considered as definitory rules. They do not tell what inferences one should draw from the available premises or what inferences people in fact draw from them always or usually. They are neither descriptive nor prescriptive rules; they are permissive ones. They tell what one can do without committing a logical mistake. They tell one how to avoid logical mistakes, known as fallacies. The study of fallacies has been part of logic ever since Aristotle. In so far as fallacies are violations of rules of inference in the narrow sense, the study of fallacies is not an independent part of logical studies. Many of the traditional fallacies are in fact not mistakes in applying deductive rules of inference, but either violations of the rules of other parts of logic or else strategic mistakes. For instance, the fallacy of begging the question did not originally mean circular reasoning, as it is nowadays viewed, but the mistake of asking directly the principal question of an interrogative inquiry. This question is supposed to be answered by putting a number of "small" questions to suitable sources of information.

This definitory vs. strategic distinction provides an interesting perspective on logical studies. In logical theorizing, a lion's share of attention has been devoted to definitory rules at the expense of strategic rules, even though the latter ones are what defines good reasoning. One historical reason for this may be the crisis in the foundations of mathematics in the early twentieth century which prompted an emphasis on the soundness of rules of logical inference. Yet for actual applications of logic strategic rules are incomparably more important.

There also prevails a confusion as to how certain types of human reasoning should be modeled by means of a logical system. Are the principles of such human reasoning to be captured by means of definitory rules or of strategic rules of reasoning? The latter answer is the more promising one, even though many developments in logic are apparently predicated on the former answer. For instance, it is fairly obvious that the principles of rational belief change should be captured by strategic rules rather than definitory rules, contrary to what many theorists are trying to do.

A characteristic difference between the two kinds of rules is that definitory rules normally concern particular moves, whereas strategic evaluation pertains in the last analysis to entire strategies in the sense of game theories. (Such strategies prescribe what a player should do in all the possible situations that he or she or it might encounter in a play of the game.)

These observations help us to understand what can correctly be meant by logical necessity. Contrary to what the dominating view of most philosophers was for a long time, there is no necessity about actually drawing logical inferences. Even when G is a logical consequence of F, one does not have to think of G as soon as one thinks of F. What is the case is that G must be the case as soon as F is true. You cannot bring it about that F without ipso facto bringing it about that G. Hence it is permissible in truth-preserving reasoning to move from F to G.

The definitory vs. strategic distinction also throws some light on the idea of nonmonotonic reasoning. The aim of logical reasoning manifesting itself in the form of a series of inferences, is normally a true ultimate conclusion (if for a moment we look away from probabilistic inferences). Hence it might seem as if the very idea of logical inferences that do not always preserve truth were an oxymoron. A solution to this problem is to point out that a sequence of inferences may ultimately lead to a true ultimate conclusion even if some of the individual inferences in the sequence are not truth preserving. This observation can be taken to vindicate the possibility of non-monotonic logics. At the same time it shows that a satisfactory theory of nonmonotonic logic must contain a strategic component, for rules of individual inferences cannot alone guarantee the truth or the probability of the conclusion of an entire sequence of nonmonotonic inferences. Such a strategic component is nevertheless missing from most nonmonotonic logics.

10 WHAT IS FIRST-ORDER LOGIC?

One way of trying to answer the title question of this article is to examine what has generally been taken to be the core area of logic. This is the logic of propositional connectives and quantifiers, known as first-order logic, predicate logic or quantification theory. (Cf. [Smullyan, 1968].) Since these are logical notions par excellence, in the light of the distinctions just made, this core area should be the study of deductive logic based on the meaning of logical concepts, studied primarily from the vantage point of correctness (definitory rules). Now the central logical concepts are generally recognized to be the (standard) quantifiers, that is, the notions expressed by “all” and “some” and propositional connectives like “and”, “or”, “if-then”, and “if and only if”. Some of the features of the behavior of quantifiers were studied already in the Aristotelian syllogistic which dominated logic until the nineteenth century. Gottlob Frege and C.S. Peirce, by removing the restriction to one-place predicates as distinguished from relations, overcame many of the limitations of syllogistic logic. The result, codified in the *Principia Mathematica* (1910-13) by Bertrand Russell and A.N. Whitehead, is a general theory of quantifiers and propositional connectives. Slowly, this theory was divided into first-order logic, in which the values of quantified variables are individuals (particulars) and higher-order logic, in which they can be entities of a higher logical type, such as sets, properties and relations. It is this first-order logic that is generally considered as the core area of logic.

What first-order logic is like is most directly explained by identifying the structures that can be discussed by its means. In an application of first-order logic, we are given a class of individuals, known variously as the domain of the model or its universe of discourse. In the inference rules for first-order logic, they are thought of as particular objects. They are represented by individual constants (names of individuals). On that domain, a number of properties and relations are defined, represented by predicates of one or more argument places. The central notions are the two quantifiers, viz. the existential quantifier $(\exists x)$ and the universal quantifier $(\forall y)$, each of which has a variable attached to it facilitating a later reference to the same individual. They express the notions of *at least one* and *every* as applied to the domain of individuals. Furthermore, the logical notation includes a number of propositional connectives, for instance negation $\sim$, conjunction &, disjunction $\vee$, and conditional ("if – then") $\supset$. In classical logic, they are assumed to be characterized by their truth-tables. Atomic formulas are formed by inserting individual constants and/or individual variables into the argument-places of predicates. The rest of the formulas are formed by repeated uses of quantifiers and propositional connectives. Often, the notion of identity is also included, expressed by $=$ with individual constants and/or variables flanking it. Formulas without free variables are called sentences. They express propositions about the universe of discourse.

First-order logic has a regular model theory. Among its metatheoretical properties there are compactness, the Löwenheim-Skolem properties, and separation properties. Most importantly, first-order logic admits of a semantically complete axiomatization, as was first shown by Gödel in 1930. These features of first-order logic are sometimes taken to be characteristic of logical systems in general. First-order logic was often considered by earlier philosophers as *the* logic, and even later their idea of what logic is, is modeled to a considerable extent on first-order logic.

11 PRESUPPOSITIONS OF THE RECEIVED FIRST-ORDER LOGIC

The central role of first-order logic is illustrated by the fact that it is widely used as a medium of representing the semantical (logical) form of natural-language sentences, including philosophical theses. Even Chomsky at one time [1986] considered his LF structures, which are the basis of semantical interpretation of English sentences, have a structure closely similar to the forms of a first-order (quantificational) formula. This practice is not without problems, however. One of the most prominent features of the notation of received first-order logic is a difference between it and the way natural language apparently operates. Natural-language verbs for being do not have a single translation into the logical notation, but have to be expressed in it differently in their different uses. (Cf. here [Hintikka, 1979].) A verb like *is* must on different occasions be treated as expressing identity, predication, or existence. On still other occasions, it may help to express class-inclusion, location or assertion (veridical *is*). Does this show that verbs like *is* are ambiguous? Only if the received notation of first-order logic is assumed to be the uniquely correct framework of semantical representation. Even though some linguists are in effect making this assumption, it is not at all obvious. It is not difficult to devise a semantic representation for English quantifiers that does not incorporate the ambiguity assumption and to formalize such a representation. In philosophy, no thinker before the nineteenth century assumed the triple ambiguity of verbs for being, that is, the alleged identity-predication-existence ambiguity.

One of the critical ingredients of first-order logic is the notion of the domain of individuals. This is a genuine novelty of modern logic. Earlier logicians admittedly used quantifiers, but for them these quantifiers were expressed by what Russell called denoting phrases (like “some Greek” or “every European”) which meant that each quantifier ranged only over a definite restricted class of values. In order to understand such a denoting phrase, aka quantifier phrase, one did not have to know what the entire universe of discourse was. This in effect made the notion of a universe of discourse redundant. Moreover, for Aristotle, the idea of a mind-independent universe of discourse would have not made any sense also because he considered realizations of forms in the soul on a par with their realizations in external reality.

Some philosophers have thought that there is only proper application of logic, viz. the actual world. For them, the determination of the basic objects of reality becomes a major philosophical problem. A more realistic view takes the applications of logic to concern sufficiently isolated parts of some actual or possible world, in analogy with probability theorists’ sample space points or physicists’ “systems”. (See [Hintikka, 2003a].) Some philosophical logicians use chunks of the actual world called “situations” as individuating applications of logic. (See [Barwise, 1981; 1989].)

The truly important novelty of first-order logic (or more generally, logic of quantifiers) does not lie in the ambiguity thesis, but in its power to express functional dependencies (dependencies between variables). Indeed, the entire first-order logic is equivalent with a quantifier-free system where all variables are universal and in which the job of existential quantifiers is done by function constants.

12 THE MEANING OF QUANTIFIERS

But is first-order logic in its received form fully representative of what logic is? The answer must turn on the meaning of the logical constants of first-order logic, especially of quantifiers and of identity. The common assumption is that quantifiers receive their meaning from their "ranging over" a class of values. An existential quantifier $(\exists x)$ prefixed to an open formula $F[x]$ says that the class of such values x satisfying $F[x]$ is not empty, and $(\forall x)F[x]$ says that all such values x satisfy $F[x]$. This assumption is among others made by Frege, who proposed that quantifiers be interpreted a higher-order predicates expressing the nonemptyness or exceptionlessness of a lower-order predicate. This is admittedly an important part of what quantifiers express. However, it is important to realize that quantifiers have another semantical role. By means of the formal dependence of quantifier of (Q_2y) on (Q_1x), we can express that the variable y depends on the variable x, in the sense of concrete factual dependence. This function of quantifiers is highly important, and it can be reduced to the "ranging over" idea only on restrictive further conditions. (See here [Hintikka, 1996; 2002a].)

But how is the formal dependence of (Q_2y) on (Q_1x) expressed? In the received logic, by the fact that (Q_2y) occurs in the syntactical scope of (Q_1x). Now in the received first-order logic these scopes are assumed to be nested. Since this nesting relation is transitive and antisymmetric, it cannot express all possible patterns of dependence and independence between variables. Hence the received first-order logic is not fully representative of the function of logical constants. This representative role is best served by the result of removing the restrictions on the patterns of formal dependence between quantifiers. The result is known as independence-friendly (IF) logic. By the same token as the received first-order logic used to be considered the core area of logic, IF logic must now be assigned that role. If so, it is in fact terminologically misleading to use any qualifying word in its name. It is the received first-order logic that deserves a special epithet because of the limitations it is subject to. In spite of this inbuilt bias, the term "IF logic" will be used in the following.

The semantics for this new basic logic is straightforward. The most central semantical concept is that of truth. A first-order sentence is true when there exist suitable "witness individuals" vouchsafing this truth. Thus e.g. $(\exists x)F[x]$ is true if and only if there exists individual b satisfying $F[x]$, and $(\forall x)(\exists y)F[x, y]$ is true if and only if for any individual a there exist b which together satisfy $F[x, y]$. As the latter example shows, such witness individuals may depend on others. The natural truth condition for a quantificational sentence S is therefore the existence of all the functions which produce these witness individuals as their values. (In the extreme case of initial existential quantifiers these functions become constants.) These functions are known as Skolem functions, and the natural definition of truth for S therefore requires the existence of a full array of Skolem functions for S. Such arrays of Skolem functions can be interpreted as winning strategies in certain games of verification and falsification called semantical games, thus throwing some light on how language-world relations are implemented. This "game-theoretical" truth definition is more elementary than Tarski-type ones.

13 SEMANTICAL GAMES

This interpretation of the all-important Skolem functions as codifying winning strategies in certain games opens interesting lines of thought concerning logic in general. The games in question are called semantical games. (See [Hintikka, 1968; Hintikka and Sandu, 1997].) The games in question are called semantical games. Does this interpretation mean that logical theory is part and parcel of the general theory of different kinds of games? The main answer is no. The reason is that the range of games relevant to the semantics of logic is fairly restricted. As an example we may consider how the information sets are determined that are presupposed in semantical games. The information set of a player making the move connected with a certain node of the game tree is the set of nodes in which the player in question can be at that time to the best of that player's knowledge. In semantical games the only important information sets are those due to the relevant player's knowledge and ignorance of what has happened at earlier moves in the same play of the game. Allowing such ignorance is the only thing needed to move from the semantics of the received first-order logic to the semantics of IF first-order logic.

Semantical games admit in fact a simple interpretation. When there is perfect information, such a game (for instance, the game $G(S)$ connected with a sentence S) can be thought of as a kind of game of verification and falsification, albeit not in the most literal sense of the expressions. From the point of the verifier, it is an attempt to find some of the "witness individuals" that would verify S. The falsifier tries to make this task as hard as possible or even impossible. From this idea, the game rules can be gathered without any difficulty. For instance, the first move in $G((\exists x)F[x])$ is the verifier's choice of an individual a from the domain. The game is then continued as in $G(F[a])$. The first move in $G((S_1\&S_2))$ is the falsifier's choice of S_1 or S_2, say S_o. The game is then continued as in $G(S_o)$.

Speaking of the verifier's choice of an individual is somewhat misleading here, for in an actual play of a game a successful move would require possibly quite elaborate searching for a right individual. Hence semantical games can also be considered as games of seeking a finding.

In spite of the natural structure of semantical games they are not devoid of subtleties. The verifier wins a play of a semantical game if and only if the play ends with a true atomic sentence. The falsifier wins if it ends up with a false one. However, the truth of a sentence cannot be defined as the verifier's win in the game $G(S)$. Rather, it must be defined as the existence of a winning strategy for the verifier. This distinction can be compared with the distinction between a sentence's being verified and being verifiable.

A philosophical perspective is obtained here by considering semantical games in the same way Wittgenstein considered his language-games, viz., as the basic semantical links between language and reality. If this idea is combined with a transcendental idea reminiscent mutatis mutandis of Kant, our knowledge of logic an be thought of as knowledge the structure and characteristics of semantical games. [Hintikka, 1973, chapter 5.]

Be this as it may, game-theoretical semantics is thus a candidate for the role of a semantical foundation of logic. In such a perspective, logic amounts to something like the study of the structure of the language-games that connect language with reality in the case of logical words.

The role of semantical games as the basic link between language and reality is also illustrated by the fact that game-theoretical semantics can be used to interpret some of the most prominent alternatives to the usual first-order logic. This can happen by restricting the set of strategies that the verifier may use, which means restricting the range of variables for Skolem functions. If they are restricted to constructive functions, we obtain an interpretation of constructivistic logic. If they were restricted to functions that are known in some reasonable sense, the result should be an interpretation of intuitionistic logic.

Game-theoretical semantics thus suggests in any case interesting perspectives relevant to the question, “What is logic?”

14 THE IMPLICATIONS OF INDEPENDENCE-FRIENDLY LOGIC: AXIOMATIZABILITY

These suggestions are reinforced and made more specific by what happens in the showcase of game-theoretical semantics, independence-friendly logic. As was pointed out, Gödel’s completeness theorem for the received first-order logic encouraged the idea that our real basic logic is semantically complete. This in turn led some thinkers to consider logic as a general study of formal systems.

If independence-friendly logic really is the authentic basic logic, such ideas must be rejected. For independence-friendly logic is not complete in the sense of there existing a formal proof procedure that in the limit produces all logical truths and only them.

At first sight, the question of axiomatizability might seem to affect only the computational implementation of different logics and more generally also of different mathematical theories. Such appearances notwithstanding, the question of complete axiomatizability makes a difference to our ideas about the nature of logic in relation to mathematics. The currently prevalent view seems to envisage an underlying logic which is semantically complete but too weak to capture all modes of mathematical reasoning. Hence the creative element in mathematics lies not in logic, but in mathematics. Furthermore, stronger principles that are needed in mathematical reasoning are typically thought of as new mathematical axioms, the first place perhaps set-theoretical axioms.

This picture is now being challenged. The alternative view suggested by independence-friendly logic is that a suitable logic can capture all the forms of mathematical reasoning. (What this suitable logic might be is discussed later in this article; see SS16–17.) This is connected with the fact that independence-friendly logic incorporates several important modes of inference that were impossible to express in the received first-order logic and likewise makes it possible to capture mathematically important concepts that would not be defined in the earlier first-order logic. These concepts include equicardinality, infinity and topological continuity. The modes of inference in question include the axiom of choice.

However, IF logic is not completely axiomatizable. Hence creativity is needed for the discovery of increasingly intricate logical truths, and no longer a monopoly of mathematicians as distinguished from logicians. This question affects our ideas about the relation of mathematics and logic and hence our ideas about logic. An important research project known as reverse mathematics is calculated to uncover the set-theoretical assumptions on which different mathematical results are based. (See [Friedman, 1970].) However, the main thrust of this project has been to ask which sets have to exist for a mathematical proof to be able to go through rather than to ask directly what stronger set-theoretical axioms are needed in the proofs. Now the existence of a set implies the possibility of applying the logical principle of tertium non datur to its definiens. Hence the project of reverse mathematics can also be viewed as a study of what applications of logical principles like tertium non datur are needed in actual mathematical proofs.

15 NEGATION AND ITS SIGNIFICANCE

Another feature of independence-friendly logic has likewise important implications for our ideas about logic in general. This feature is the behavior of negation. (See here [Hintikka, forthcoming (b)].) The semantical game rule (or rules) for negation in independence-friendly logic are the same as in the received logic of quantifiers. Yet the resulting negation turns out not to be the contradictory negation that is often thought of as being *the* negation. Instead, we receive a stronger dual negation that does not obey the law of excluded middle.

From a game-theoretical viewpoint this is only to be expected, for the law of excluded middle amounts to assuming that semantical games are determinate. And of course there are plenty of perfectly natural games that are not determinate.

Many philosophers will react to this by saying that what it shows is that independence friendly logic is a “nonclassical” logic. But what is their criterion of a logic’s being classical? In view of the “classical” character of the game rules for negation in independence-friendly logic, it appears quite as natural to maintain that this logic shows that the law of excluded middle is not part and parcel of the classical conception of logic.

The absence of tertium non datur means that independence-friendly logic is in some ways closer to intuitionistic logic than the received first-order logic, in spite of its greater expressive power.

Of course we have to introduce also a contradictory negation $\neg$ into the independence-friendly logic over and above the dual negation $\sim$ in order to reach an adequate basic logic. The result will be called the extended independence-friendly first-order logic. The contradictory negation cannot be characterized by game rules however, for the “classical” ones yield the dual one. Hence $\neg$ can occur in extended independence-friendly logic only sentence-initially.

The extended independence-friendly logic has an interesting structure. Algebraically, it has the structure of a Boolean algebra with an operator in Tarski's sense. (See [Jónsson and Tarski, 1951; 1952].) In a sense, this structure is therefore the true algebra of logic. By Tarski's results, the extended independence logic admits to a set-theoretical ("geometrical") interpretation. In this interpretation the strong negation turns out to express a generalization of the notion of orthogonality. In terms of orthogonality, we can even define such notions as dimensions, coordinate representation etc. for purely logical spaces. (See [Hintikka, 2004c; forthcoming (b)].)

In sufficiently strong languages, there must thus be two different notions of negation present, the strong dual negation and a contradictory one. This presumably applied also to natural languages even though in them there usually is but one grammatical construction for negation. This puts the notion of negation in natural languages to a new light.

If we make suitable further assumptions concerning the logical spaces with two negations, we obtain more elaborate algebraic structures which largely remain to be examined. By using game-theoretical ideas, we can also introduce in a natural way probability measures on logical spaces. Whether, and if so in what sense, this makes probability a logical concept remains to be examined.

16 HIGHER-ORDER LOGICS

These developments throw new light also on second-order logic and other higher-order logics. First, however, it is in order to see what these logics are like.

First-order logic is characterized by the fact that all quantifiers range over all the individuals in the given universe of discourse. Second-order logic is obtained when quantifiers are admitted that range over sets and other predicates-in-extension of individuals. They are known as second-order quantifiers and their values second-order entities. Third-order quantifiers range over sets and predicates-in-extensions of second-order entities, and so on. Instead of such extensional entities, the concepts that define them are sometimes employed in higher-order logics as values of higher-order variables. We can thus distinguish extensional and intensional interpretations of higher-order logics. When the different orders are separated from each other, the result is also known as a theory of types in a broad sense.

Within extensionally construed higher-order logics we can distinguish two different interpretations. If the higher-order quantifiers range over all extensionally possible entities of the appropriate lower type, we obtain what is known as the standard interpretations. If they range over some designated subset of such entities, we are dealing with a nonstandard interpretation. The most prominent such nonstandard interpretation is the one in which higher-order quantifiers range over entities definable in the language that is being used. Often, the term "non standard interpretation" is restricted to this one.

The distinction was first formulated and the terms "standard" and "nonstandard" introduced by Leon Henkin in 1950. However, the distinction was clear to Frank

Ramsey who proposed transforming the intensional higher-order system of Russell's and Whitehead's *Principia Mathematica* into a form in which it can be given a standard interpretation. (See [Ramsey, 1925].)

Even earlier, a special case of the standardness idea had played an important role in the foundations of mathematics in the form of the notion of an arbitrary function. The class of all such functions is nothing but the standard range of the function variables of the appropriate type. The important development in the history of mathematics of the idea of an arbitrary function is in effect the development of the standard interpretation of function variables. Applied to functions from natural numbers to integers the notion of arbitrary function becomes the idea of an arbitrary sequence of integers that has played an important role in the general theory of functions, especially in their representations as series of different kinds.

For most mathematical conceptualizations and modes of reasoning, second-order logic with standard interpretation is amply sufficient. Such a logic is not semantically complete, however.

Higher-order logics are in a certain sense parallel to set theory. An axiomatized set theory is like a higher-order logic without type restrictions. The logical principles that are normally used in set theory are those of first-order logic, not of higher-order logic. The question of the adequacy of such first-order axiomatizations of set theory must still be considered as being open. (See [Hintikka, forthcoming (a)].)

The difficulties in this direction are in the first instance due to the fact that the standard interpretation cannot be imposed on the models of a system of set theory by first-order axioms. Hence the idea of a model of an axiomatic first-order set theory can be understood in two different ways. On one interpretation, the would-be membership relation $\in$ is interpreted like any first-order predicate. But we can also require that it be actually interpreted as the membership relation. One source of difficulties here is revealed by the question whether an axiomatic set theory can have models that can be interpreted in the latter way even locally.

The question of the existence of higher-order entities depends on their mode of existence and on their knowability (or lack thereof). Several philosophers have been suspicious of such entities and would rather dispense with them altogether. W.V. Quine among others has for this reason preferred set theory to higher-order logic. This is a questionable position, however, for an attempt to interpret the first-order quantifiers of an axiomatized set theory set-theoretically leads to serious difficulties. (See [Hintikka, 2004a; forthcoming (a)].)

In any case, in either approach assumptions are needed that go beyond first-order logic, either higher-order principles of reasoning or axioms of set existence. A typical example is the axiom of choice which unsurprisingly has been a bone of contention in the foundations of mathematics. Hilbert's epsilon-calculus can be viewed as a large-scale attempt to reduce the axiom of choice to the first-order level. (See [Hilbert and Bernays, 1934–1939].)

17 A DISTINCTION WITHOUT ONTOLOGICAL DIFFERENCE?

Certain recent advances in logic have disturbed the neat distinction between first-order and higher-order logic, however. (See [Hintikka, forthcoming (b)].) For one thing, independence-friendly logic captures on the first order level several important concepts and modes of reasoning that earlier were thought of as being possible to capture only by second-order means. As was noted they include the notions of equicardinality and infinity, the axiom of choice, König's lemma and topological continuity. In the extended independence-friendly logic we can add to the list among others the notions of well-ordering and mathematical induction. Their availability means that we can on the first-order level (that is, without quantifying over higher-order entities) carry out enormously more reasoning, including mathematical reasoning, than was previously thought of as being possible to do on this level. Moreover, this is possible not only without quantifying over higher-order entities but also without evoking the principle of excluded middle, in the sense that this principle is not assumed in independence-friendly logic. However, there are modes of logical reasoning that cannot be so captured and propositions that cannot be expressed in extended independence-friendly logic. The prime examples are offered by prima facie propositions in which the contradictory negation $\neg$ occurs within the scope of a quantifier. Such formulas cannot be assigned an interpretation by means of game-theoretical semantics. They can only be interpreted by reference to the totality of substitution-instances of their non-quantified part, the substitution-values being of course the names of all the members of the domain. When this domain is infinite, this involves an application of tertium non datur to infinite sets as closed totalities. The resulting meanings are as a consequence nonelementary ("infinitistic") in a sense in which independence-friendly logic is not. (Semantical games do not involve infinite sets as closed totalities.)

At the same time the resulting infinitistic logic is first-order in the sense that no quantification over second-order or higher-order entities is involved. It nevertheless turns out that the resulting logic — that is, extended independence-friendly logic reinforced by contradictory negation which is allowed to occur within the scope of quantifiers and moreover allowed to be arbitrarily nested — is as strong as the entire second-order logic. Since second-order logic with standard interpretation is as strong a logic as anybody is likely to need in science or classical mathematics, it is thus seen that all logic we reasonably may need can be done ontologically speaking on the first-order level. Quine has said that to be is to be a value of a bound variable. If that is the case, we don't need the existence of any second-order (or other higher-order) entities. In any case it is true that we need not worry about what the existence of higher-order entities means or which of them must be assumed to exist.

Instead, what distinguishes the kind of reasoning that is now discussed in the context of higher-order logics from elementary reasoning is the unrestricted use of tertium non datur. Such use was in fact considered as the mark of infinitistic reasoning by the likes of David Hilbert and L.E.J. Brouwer [1923]. Whether this means that the contrast between first-order logic and higher-order logic disappears is a mere matter of terminological preference.

18 ALTERNATIVE LOGICS?

All this leaves in the dark the status of the many “alternative logics” that have sprung up in recent years. (See [van Benthem, forthcoming; Ginsberg 1987].) Since they are typically non-monotonic, they cannot be truth-preserving or deductive. But if so, how can they be claimed to be alternative to the traditional systematizations of deductive logic? And if they embody ampliative inference, what is the new information that is brought in and where does it come from? And if they are (as inductive logic was argued to be) dependent on unspoken assumptions, what are those assumptions?

In order to answer such questions, we can consider as an example one of the most clear-cut “alternative logics”, developed initially by John McCarthy [1980] and called “inference by circumscription”. Its basic idea is clear. The inferences considered in the theory of circumscriptive inference are based not only on their explicit premises in the usual sense, but also on the tacit assumption that these premises constitute all the relevant information. It is amply clear that such inferences are common in ordinary discourse. It is also amply clear that such tacit assumptions of exhaustive relevance are not unavoidable. For instance, a typical tactic in solving run-of-the-mill recreational puzzles is to bring in information that is not contained in the given premises. However, there are no overwhelming reasons in evidence to show that the reasoning itself that is required for circumscriptive inference is different from our ordinary ways of reasoning. Hence it seems that we can simply pidgeonhole circumscriptive inference as a chapter in the theory of enthymemic reasoning, that is, in the theory of reasoning from partly tacit premises. And if so, circumscriptive inference can scarcely be claimed to be alternative to our normal logical inferences — or so it seems.

However, this is not the end of the story. The most important peculiarity of the tacit circumscriptive assumption is that it is not expressible in the language in which the inferences are carried out. This language is normally some familiar logical one not unlike a first-order language. However, the exhaustiveness of the given premises obviously cannot be expressed in such a language.

This shows both the interest of circumscriptive inference and the equivocal status of an attempt to formulate such inferences on a first-order level. The theoretically satisfactory approach would be to develop a logical language in which the tacit premise can be explicitly formulated. Only then could it be decided whether inferences from such premises require a logic different from our well-known modes of inference. It is also reasonable to expect that the logic of such richer languages would be an extension of, rather than an alternative to, the established logics

One can in fact view many other kinds of ampliative reasoning in the same way. This provides an interesting perspective on a number of “alternative logics”. And this perspective is not restricted to recently developed “alternative logics”.

The hidden assumptions on which applications of inductive logic can be argued to depend cannot be formulated in first-order languages even after simple probability measures are added to it. The theoretical situation hence seems to be the same as in the logic of circumscription, even though no one has tried to label inductive logic an "alternative" logic.

Similar situations occur also in mathematics. An interesting example is offered by maximality assumptions. Hilbert tried to make his axiomatization of geometry categorical by means of a maximality assumption that he called the axiom of completeness. ([Hilbert, 1899]; this axiom was actually introduced only in the second edition.) Gödel surmised that maximality assumptions not unlike Hilbert's axiom are what is needed in axiomatic set theory. Such assumptions are nevertheless hard to express in the usual logical languages like first-order languages. The reason is that the propositions of such a language work by imposing conditions on their models one model at the time, whereas maximality assumptions involve comparisons of different models with each other. This situation differs from the alternative logics in that Hilbert tried to formulate the additional axiom explicitly. However, such a formulation is impossible on the level on which most of the other reasoning is in Hilbert's book.

One can perhaps view first-order axiomatic set theory in the same light. The logic used there is precisely the traditional first-order logic. However, the characteristically set-theoretical modes of inference, such as the axiom of choice, cannot be adequately captured in traditional first-order logic. Hence the set theory incorporates only *Ersatz* versions of these principles of inference in the forms of axioms.

Why have logicians not developed richer languages in which such tacit assumptions could be expressed explicitly? Part of an answer was in the difficulty of such an enterprise. But perhaps part of the reason is that logicians have entertained too narrow a conception of what logic is. For one cannot help suspecting that somehow logicians have assumed that the familiar languages like first-order languages are capable of expressing all logical inferences. If so, we have here a telling example of the hegemony of traditional first-order logic as an answer to the title question of this article. Even logics that are calculated to be alternatives to the traditional first-order logic have been unwittingly colored by a belief in its unique role.

It nevertheless appears that the right project is to develop richer logical languages rather than devise different sets of special-interest modes of reasoning. If and when this is done, the result is likely to be a richer idea of what logic is – or of what it can be.

19 AMPLIATIVE INFERENCE AND ACQUISITION OF INFORMATION

In a different direction one can still ask for a unified perspective on ampliative inference. Such an overview on nondeductive inference must be connected with the notion of information, for the greatest common denominator of all ampliative inferences is that new information enters into an ongoing argument. A general theory of such inferences will have to focus on the source of the new information in comparison with alternative sources. For one thing, the reliability of the new information depends on the reliability of its source. Likewise, the evaluation of a new item of information depends on a comparison with the other items of information that might have been available instead of the actually introduced one.

One way of trying to reach such a general theory is to consider the reception of any new items of information as an answer to a (usually tacit) question. Indeed, if we know the source of an item of information and know what other items might be available from the same source or from others, etc., we conceptually speaking might as well consider such acquisition of the information as the reception of an answer to a question. From such a point of view, the theory of questions and answers assumes the role of a general framework for different kinds of ampliative inference. (See [Hintikka, 1999; Hintikka, Halonen and Mutanen, 1999].)

The theory of questions and answers is in turn based on epistemic logic, for a question is essentially a request of information [Hintikka, 1976; 2003b]. As was indicated in §3 above, a suitable interrogative logic can be developed as being determined largely as the logic of the declarative propositions that specify the epistemic state of affairs that a questioner asks to be brought about. Such a proposition is called the desideratum of the question in question. Needless to say, both the notion of a question and the idea of a source of answers must then be taken in a very general sense. The answerer need not be a human being or a database. It can be nature, one's environment, memory or even a questioner's imagination. Likewise, the act of asking a question may for instance take the form of an experiment or observation, or perhaps even a guess.

Such an "interrogative model of inquiry", as it has been called, is especially useful in the study of the strategic aspects of information seeking. For such a search amounts to a question-answer sequence interspersed naturally by deductive inferences. Strategies of inquiry will then amount to different methods of choosing the questions. The possibilities of theorizing opened by this perspective have not yet been used very widely. It is nevertheless clear that for instance some of the traditional "fallacies" are not breaches of the definitory rules of inference but violations of the reasonable principle of questioning, among them the so-called fallacy of petitio principii. (Cf. [Hamblin, 1970; Hintikka, 1987].)

Such interrogative inquiry is more closely related to the theory of the deductive inference than perhaps first meets the eye. This relationship in fact throws interesting light on the nature of logic in general. Formally, interrogative inquiry is partly analogous to deductive inquiry. In order to be in a position to ask a question, an inquirer must have established its presupposition. A step from the presupposition of a question to its answer (whenever available) is formally similar to a step from the premise (or premises) of a deductive inference to its conclusion. As a consequence, strategies of questioning govern the choices of presuppositions from the set of available propositions while strategies of deduction govern the choices of premises from the same pool of propositions.

In neither case can the optimal strategies be expressed in the form of recursive (mechanical) rules. However, there is a deep connection between the two kinds of strategies. In a context of pure discovery, that is, when all available answers are true and known to be true, the optimal strategies of interrogative inquiry coincide with the optimal strategies of deductions except for deviations caused by the possible unavailability of some answers. (Actually, another discrepancy may arise when the answers happen to deal with previously known entities.)

In other (somewhat looser) words, in a context of pure discovery interrogative inquiry is guided in effect by the same strategies as deduction. This result throws important light on what logic in general is, in particular what the place of deductive logic is in a wider schema of epistemological ideas. It suggests looking at deductive logic as a kind of systematic study *a priori* of what can happen in an actual empirical enterprise of information acquisition. It also goes a long way toward vindicating the old and still "Sherlock Holmes view" of logic as the gist of all good reasoning mentioned in §1 above. More generally speaking, it shows the relationship between deductive and ampliative reasoning.

BIBLIOGRAPHY

Comprehensive surveys of different aspects of logic are Jon Barwise, editor, 1977, *Handbook of Mathematical Logic*, North-Holland, Amsterdam; D. Gabbay and F. Guenther (eds.), *Handbook of Philosophical Logic*, second ed., Kluwer Academic, Dordrecht, 2004; and D. Gabbay, C.J. Hogger and J.A. Robinson (eds.), 1993–1996, *Handbook of Logic in Artificial Intelligence and Logic Programming* 1-5, Oxford University Press, Oxford.
Other literature listed here pertains to particular subjects covered in this article even though not all of them are referred to in so many words in the text. They are not intended to exhaust the subject, only to provide initial clues to the relevant literature.

[Anderson *et al.*, 1992] A. Anderson, N. Belnap and M. Dunn. *Entailment: The Logic of Relevance and Necessity II.* Princeton: Princeton University Press, 1992.

[Antoniou, 1997] G. Antoniou. *Nonmonotonic Reasoning.* Cambridge, Massachusetts: The MIT Press, 1997.

[Barwise, 1981] J. Barwise. Scenes and other situations. *Journal of Philosophy*, 77: 369-397, 1981.

[Barwise, 1989] J. Barwise. *The Situation in Logic.* Stanford: CSLI, 1989.

[Beth, 1947] E.W. Beth. Hundred years of symbolic logic: a retrospect on the occasion of the Boole-De Morgan centenary. *Dialectica*, 1: 331-345, 1947.

[Beth, 1953] E.W. Beth. On Padoa's method in the theory of definition. *Nederl. Akad. Wetensch. Proc. Ser. A*, 56: 330-339, 1953.

[Beth, 1962] E.W. Beth. *Formal Methods.* Dordrecht: D. Reidel, 1962.

[Bishop, 1967] E. Bishop. *Elements of Constructive Analysis.* New York: McGraw-Hill, 1967.

[Blackburn *et al.*, 2001] P. Blackburn, M. de Rijke and Y. Venema. *Modal Logic.* Cambridge, 2001.

[Boole, 1958] G. Boole. *The Laws of Thought.* New York: Dover, 1958 (original 1854).

[Boole, 1997] G. Boole. *Selected Manuscripts on Logic and its Philosophy.* I. Grattan-Guiness and G. Bornet (eds.), Basel: Birkhäuser Verlag, 1997.

[Boolos, 1975] G. Boolos. On second-order logic. *Journal of Philosophy*, 72: 509-527, 1975.

[Boolos and Jefffrey, 1989] G. Boolos and R. Jeffrey. *Computability and logic.* Third edition, Cambridge, England: Cambridge University Press, 1989.

[Brower, 1923] L.E.J. Brouwer. On the significance of the principle of excluded middle in mathematics (1923). In J. van Heijenoort (ed.), *From Frege to Gödel*, Cambridge, Massachusetts: Harvard University Press, 334-345, 1967.

[Buss, 1998] S. Buss (ed.). *Handbook of Proof Theory.* Amsterdam and New York: Elsevier, 1998.

[Carnap, 1947] R. Carnap. *Meaning and Necessity.* Chicago: University of Chicago Press, 1947.

[Carnap, 1950] R. Carnap. *Logical Foundations of Probability.* Chicago: University of Chicago Press, 1950.

[Carnap, 1952] R. Carnap. *The Continuum of Inductive Methods.* Chicago: University of Chicago Press, 1952.

[Chang and Keisler, 1990] C.C. Chang and J. Keisler. *Model Theory.* Amsterdam: North-Holland, 1990.

[Chomsky, 1975] N. Chomsky. *The Logical Structure of Linguistic Theory.* New York: Plenum Press, 1975.

[Chomsky, 1986] N. Chomsky. *Knowledge of Language.* New York: Praeger, 1986.

[Church, 1956] A. Church. *Introduction to Mathematical Logic.* Princeton: Princeton University Press, 1956.

[Chang and Keisler, 1991] J.A. Coffa. *The Semantic Tradition from Kant to Carnap.* L. Wessels (ed.), Cambridge, Massachusetts: Cambridge University Press, 1991.

[Davis, 1965] M. Davis (ed.). *The Undecidable.* Hewlett, N.Y.: Raven Press, 1965.

[Dreben and Floyd, 1991] B. Dreben and J. Floyd. Tautology: How not to use the word. *Synthese*, 87: 23-49, 1991.

[Dummett, 1977] M. Dummett. *Elements of Intuitionism.* Oxford: Clarendon Press, 1977.

[Ebbinghaus and Flum, 1999] H.D. Ebbinghaus and J. Flum. *Finite Model Theory.* Second edition, New York: Springer, 1999.

[Etchemendy, 1983] J. Etchemendy. The doctrine of logic as form. *Linguistics and Philosophy*, 6: 319-334, 1983.

[Etchemendy, 1990] J. Etchemendy. *The Concept of Logical Consequence.* Cambridge, Masschusetts: Harvard University Press, 1990.

[Fine, 1985] K. Fine. *Reasoning with Arbitrary Objects.* Oxford: Blackwell, 1985.

[Fine, 1995] K. Fine. Part-whole. In B. Smith and D.W. Smith (eds.), *The Cambridge Companion to Husserl*, Cambridge, England: Cambridge University Press, 463-485, 1995.

[Frege, 1879] G. Frege. *Begriffschrift* (1879). In J. van Heijenoort (ed.), *From Frege to Gödel*, Cambridge, Massachusetts: Harvard University Press 1967.

[Frege, 1970] G. Frege. *Posthumous Writings.* Chicago: University of Chicago Press, 1970.

[Friedman, 1970] H. Friedman. Higher set theory and mathematical practice. *Annals of Mathematical Logic*, 2: 325-357, 1970.

[Gabbay, 1976] D. Gabbay. *Investigations in Modal and Tense Logics.* Dordrecht: D. Reidel, 1976.

[Gabbay *et al.*, 1993–1996] D. Gabbay, C. J. Hogger and J. A. Robinson, eds. *Handbook of Logic in Artificial Intelligence and Logic Programming*, 5 Volumes. Oxford: Clarendon Press, 1993–1996.

[Gabbay, 1994] D. Gabbay (ed.). *What is a Logical System?* Oxford: Clarendon Press, 1994.

[Gabbay and Guenthner, 2001-2005] D. Gabbay and F. Guenthner (eds.) *Handbook of Philosophical Logic* (13 volumes), New York: Springer, 2001- 2005.

[Ginsberg, 1987] M. Ginsberg (ed.). *Readings in Nonmonotonic Reasoning.* Los Altos, California: Morgan Kauffman, 1987.

[Goldfarb, 1979] W. Goldfarb. Logic in the twenties, the nature of a quantifier. *Journal of Symbolic Logic*, 44: 351-368, 1979.

[Gödel, 1930] K. Gödel. The completeness of the axioms of the functional calculus of logic (1930). In J. van Heijenoort (ed.), *From Frege to Gödel*, Cambridge, Massachusetts: Harvard University Press, 582-591, 1967.

[Gurevich and Harrington, 1984] Y. Gurevich and L. Harrington. Trees, automata, and games. In H.R. Lewis (ed.), *Proceedings of the ACM Symposium on the Theory of Computing*, ACM, San Francisco, 171-182, 1984.

[Hacking, 1979] I. Hacking. What is logic? *Journal of Philosophy*, 76: 285-319, 1979.

[Halpern, 2003] J.Y. Halpern. *Reasoning About Uncertainty.* Cambridge, Massachusetts: The MIT Press, 2003.

[Hamblin, 1970] C. Hamblin. *Fallacies.* London: Methuen, 1970.

[Hansen and Pinto, 1995] H.V. Hansen and R.C. Pinto (eds.). *Fallacies: Classical and Contemporary Readings*, University Park, Pennsylvania: Penn State Press, 1995.

[Hempel, 1945] C.G. Hempel. On the nature of mathematical truth. *American Mathematical Monthly*, 24: 543-556, 1945.

[Henkin, 1950] L. Henkin. Completeness in the theory of types. *Journal of Symbolic Logic*, 15: 81-91, 1950.

[Heyting, 1956] A. Heyting. *Intuitionism: An Introduction*. Amsterdam: North-Holland, 1956.

[Hilbert, 1899] D. Hilbert. *Grundlagen der Geomatrie*. Teubner, Leipzig, 1899. (Second ed., 1902.)

[Hilbert, 1918] D. Hilbert. Axiomatisches Denken. *Mathematsche Annalen*, 78: 405-420, 1918.

[Hilbert, 1926] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95: 161-190, 1926.

[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*. Berlin: Springer, 1928.

[Hilbert and Bernays, 1934-1939] D. Hilbert and P. Bernays. *Grundlagen der Mathematik* 1-2, Heidelberg: Springer, 1934-1939.

[Hilpinen, 1971] R. Hilpinen. *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: D. Reidel, 1971.

[Hintikka, 1962] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, New York: Cornell University Press, 1962.

[Hintikka, 1965] J. Hintikka. Are logical truths analytic? *Philosophical Review*, 74: 178-203, 1965.

[Hintikka, 1968] J. Hintikka. Language-games for quantifiers. In N. Rescher (ed.), *Studies in Logical Theory*, Oxford: Blackwell, Oxford, 46-73, 1968.

[Hintikka, 1970] J. Hintikka. Depth information vs. surface information. In J. Hintikka and P. Suppes (eds.), *Information and Inference*, Dordrecht: D. Reidel, 263-297, 1970.

[Hintikka, 1973] J. Hintikka. *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*. Oxford: Clarendon Press, 1973.

[Hintikka, 1976a] J. Hintikka. A new approach to infinitary languages. *Annals of Mathematical Logic*, 10: 95-115, 1976a.

[Hintikka, 1976b] J. Hintikka. *The Semantics of Questions and the Questions of Semantics*. (*Acta Philosophica Fennica*, 28(4), Amsterdam: North-Holland, 1976b.

[Hintikka, 1979] J. Hintikka. 'Is', semantical games and semantical relativity. *Journal of Philosophical Logic*, 8: 433-468.

[Hintikka, 1983] J. Hintikka. Semantical games, the alleged ambiguity of 'is', and Aristotelian categories. *Synthese*, 54: 443-469, 1983.

[Hintikka, 1987] J. Hintikka. The fallacy of fallacies. *Argumentation*, 1: 211-238, 1987.

[Hintikka, 1993] J. Hintikka. The concept of induction in the light of the interrogative approach to inquiry. In J. Earman (ed.), *Inference, Explanation and other Frustrations*, Berkeley: University of California Press, 23-43, 1993.

[Hintikka, 1995] J. Hintikka. The games of logic and the games of inquiry. *Dialectica*, 49: 229-249, 1995.

[Hintikka, 1997] J. Hintikka. No scope for scope? *Linguistics and Philosophy*, 20: 515-544, 1997.

[Hintikka, 1999a] J. Hintikka. *Inquiry as Inquiry: A Logic of Scientific Discovery* (*Selected Papers* vol.5). Dordrecht: Kluwer Academic, 1999a.

[Hintikka, 1999b] J. Hintikka. Is the axiom of choice a logical or a set-theoretical principle? *Dialectica*, 53: 283-290, 1999b.

[Hintikka, 2001] J. Hintikka. Is logic the key to all good reasoning? *Argumentation*, vol.15: 35-57, 2001.

[Hintikka, 2002a] J. Hintikka. Hyperclassical Logic (a.k.a IF Logic) and its implications for logical theory. *Bulletin of Symbolic Logic*, 8: 404-423, 2002a.

[Hintikka, 2002b] J. Hintikka. Quantum logic as a fragment of independence-friendly logic. *Journal of Philosophical Logic*, 31: 197-209, 2002b.

[Hintikka, 2003a] J. Hintikka. A distinction too few or too many? In C. Gould (ed.), *Constructivism and Practice*, Lanham, Maryland: Rowman & Littlefield, 47-74, 2003a.

[Hintikka, 2003b] J. Hintikka. A second-generation epistemic logic. In V. Hendricks *et al.* (eds.), *Knowledge Contributors*, Dordrecht: Kluwer Academic, 33-55, 2003b.

[Hintikka, 2004a] J. Hintikka. Independence-friendly logic and axiomatic set theory. *Annals of Pure and Applied Logic*, 126: 313-333, 2004a.

[Hintikka, 2004b] J. Hintikka. Logical versus nonlogical concepts: An untenable dualism? In S. Rahman *et al.* (eds.), *Logic, Epistemology and the Unity of Science*, Dordrecht: Kluwer Academic, 51-56, 2004b.

[Hintikka, 2004c] J. Hintikka. What is the true algebra of logic? In V. Hendricks *et al.* (eds.), *First-Order Logic Revisited*, Logos, Berlin, 117-128, 2004c.

[Hintikka, forthcoming (a)] J. Hintikka. Truth, axiom of choice and set theory: what Tarski never told us. Forthcoming.

[Hintikka, forthcoming (b)] J. Hintikka. Truth, negation, and other basic notions of logic. In J. van Benthem (ed.), *The Age of Alternative Logics*. Forthcoming.

[Hintikka, forthcoming (c)] J. Hintikka. Who has kidnapped the notion of information? Forthcoming.

[Hintikka *et al.*, 1999] J. Hintikka, I. Halonen and A. Mutanen. Interrogative logic as a general theory of reasoning. In D. Gabbay *et al.* (eds.), *Studies in Logic and Practical Reasoning*, Dordrecht: Kluwer Academic, 1999.

[Hintikka and Hintikka, 1982] J. Hintikka and M. Hintikka. Sherlock Holmes confronts modern logic: Toward a theory of information-seeking through questioning. In E.M. Barth and J.L. Martens, J. Benjamins (eds.), *Argumentation: Approaches to Theory Formation*, Amsterdam, 55-76, 1982.

[Hintikka and Sandu, 1997] J. Hintikka and G. Sandu. Game-theoretical semantics. In J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Amsterdam, Elsevier, 361-410, 1997.

[Hintikka and Suppes, 1970] J. Hintikka and P. Suppes (eds.), *Information and inference*, Dordrecht: D. Reidel, 1970.

[Hodges, 1985] W. Hodges. *Building Models by Games.* Cambridge: Cambridge University Press, 1985.

[Hodges, 1993] W. Hodges. *Model Theory.* Cambridge: Cambridge University Press, 1993.

[Hodges, 2001] W. Hodges. Elementary Predicate Logic 25: Skolem Functions. In D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic I*, second edition, Dordrecht: Kluwer Academic, 86-91, 2001.

[Hooker, 1975-1977] C.A. Hooker (ed.), *The Logico-Algebraic Approach to Quantum Mechanics* 1-2, D, Dordrecht: D. Reidel, 1975-1977.

[Husserl, 1913] E. Husserl. *Logische Untersuchungen.* Tübingen: Max Niemeyer, 1913.

[Japaridze, 2003] G. Japaridze. Introduction to computability logic. *Annals of Pure and Applied Logic*, 123: 1-99, 2003.

[Jeffrey, 1983] R. Jeffrey. *The Logic of Decision.* Second edition, Chicago: University of Chicago Press, 1983.

[Jeffrey, 1992] R. Jeffrey. *Probability and the Art of Judgment.* New York: Cambridge University Press, 1992.

[Jónsson and Tarski, 1951] B. Jónsson and A. Tarski. Boolean algebras with operators I. *American Journal of Mathematics*, 73: 891-939, 1951. Also in Tarski 1986, vol.3, 369-419.

[Jónsson and Tarski, 1952] B. Jónsson and A. Tarski. ibid. II. *American Journal of Mathematics*, 74: 127-162, 1952. Also in Tarski 1986, vol. 3, 421-458.

[Keisler, 1974] J. Keisler. *Model Theory for Infinitary Logic.* Amsterdam: North-Holland, 1974.

[Kneale and Kneale, 1962] W. Kneale and M. Kneale. *The Development of Logic.* Oxford: Oxford University Press, 1962.

[Knnuuttila and Hintikka, 1986] S. Knnuuttila and J. Hintikka (eds.). *The Logic of Being*, Dordrecht: D. Reidel, 1986.

[Kripke, 1963] S. Kripke. Semantical Considerations on Modal Logic. *Acta Philosophica Fennica*, 16: 83-94, 1963.

[Kyburg, 1974] H. Kyburg. *The Logical Foundations of Statistical Inference.* Dordrecht: D. Reidel, 1974.

[Lemmon and Scott, 1977] E.J. Lemmon and D. Scott. *An Introduction to Modal Logic.* Oxford: Blackwell, 1977.

[Levi, 1977] I. Levi. Direct Inference. *Journal of Philosophy*, 74: 5-29, 1977.

[Lewis, 1973] D. Lewis. *Counterfactuals.* Cambridge, Massachusetts: Harvard University Press, 1973.

[Lewis, 1980] D. Lewis. A Subjectivist's Guide to Objective Chance. In R.C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, vol. 2, Berkeley: University of California Press, 263-293, 1980.

[Lycan, 1984] W. Lycan. *Logical form in natural language.* Cambridge, Massachusetts: The MIT Press, 1984.

[May, 1985] R. May. *Logical Form: Its Structure and Derivation.* Cambridge, Massachusetts: The MIT Press, 1985.

[McCarthy, 1980] J. McCarthy. Circumscription — a form of non-monotonic reasoning. *Artifical Intelligence*, 13: 27-39, 171-172, 1980.

[McCarthy, 1981] T. McCarthy. The idea of a logical constant. *Journal of Philosophy*, 78: 499-523, 1981.

[Mendelson, 1987] E. Mendelson. *Introduction to mathematical logic.* Third edition, Princeton: van Nostrand, 1987.

[Mostowski, 1957] A. Mostowski. On a generalization of quantifiers. *Fundamenta Mathematicae*, 44:12-36, 1957.

[Mostowski, 1966] A. Mostowski. *Thirty Years of Foundational Studies*, Oxford: Basil Blackwell, 1966.

[Parikh, 1985] R. Parikh. The logic of games and its applications. In M. Karpinski and J. van Leeuwen (eds.), *Annals of Discrete Mathematics*, 24: 111-140, 1985.

[Peirce, 1885] C.S. Peirce. On the Algebra of Logic; A Contribution to the Philosophy of Notation. *American Journal of Mathematics*, 7: 180 – 202, 1885.

[Peirce, 1992] C.S. Peirce. *Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898.* Kenneth Laine Ketner, editor, Cambridge, Massachusetts: Harvard University Press, 1992.

[Prawitz, 1985] D. Prawitz. Remarks on some approaches to the concept of logical consequence. *Synthese*, 62: 153–171, 1985.

[Preyer and Peters, 2002] G. Preyer and G. Peters (eds.). *Logical Form and Language.* Oxford: Oxford University Press, 2002.

[Quine, 1950] W.V.O. Quine. *Methods of Logic.* New York: Henry Holt, 1950.

[Quine, 1951] W.V.O. Quine. Two dogmas of empiricism. *Philosophical Review*, 60: 20-43, 1951.

[Quine, 1970] W.V.O. Quine. *Philosophy of Logic.* Cambridge Massachusetts, Harvard University Press, 1970.

[Ramsey, 1925] F.P. Ramsey. The foundations of mathematics. *Proceedings of the London Mathematical Society*, Ser. II, 25: 338-384, 1925.

[Ray, 1996] G. Ray. Logical Consequence: A Defense of Tarski. *The Journal of Philosophical Logic*, 25: 617-677, 1996 .

[Restall, 2000] G. Restall. *An Introduction to Substructural Logics.* Routledge, 2000.

[Sainsbury, 1991] M. Sainsbury. *Logical Forms.* Oxford: Blackwell, 1991.

[Scott, 1965] D. Scott. Logic with Denumerably Long Formulas and Finite Strings of Quantifiers. In J. Addison, L. Henkin and A. Tarski (eds.), *The Theory of Models*, Amsterdam: North-Holland, 329-341, 1965.

[Shapiro, 1998] S. Shapiro. Logical consequence, models and modality. In M. Schirn (ed.), *Philosophy of Mathematics Today*, Oxford: Clarendon Press, 131-156, 1998.

[Smith, 1982] B. Smith. *Parts and Moments: Studies in Logic and Formal Ontology.* Munich: Philosophia Verlag, 1982.

[Smullyan, 1968] R. Smullyan. *First-Order Logic.* Heidelberg and New York: Springer, 1968.

[Sher, 1991] G. Sher. *The Bounds of Logic.* Cambridge, Massachusetts: The MIT Press, 1991.

[Skyrms, 1966] B. Skyrms. *Choice and Chance: An Introduction to Inductive Logic.* Belmont California: Dickenson, 1966.

[Stalnaker, 1984] R. Stalnaker. *Inquiry.* Cambridge, Massachusetts: The MIT Press, 1984.

[Stalnaker, 1994] R. Stalnaker. Nonmonotonic consequence relations. *Fundamenta Informaticae*, 21: 7-21, 1994.

[Tarski, 1935] A. Tarski. Der Wahreitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1: 261-405, 1935.

[Tarski, 1941] A. Tarski. On the Calculus of Relations. *Journal of Symbolic Logic*, 6: 73 – 89, 1941.

[Tarski, 1944] A. Tarski. The Semantic Conception of Truth. *Philosophy and Phenomenological Research*, 4: 341-75, 1944.

[Tarski, 1956] A. Tarski. *Logic, Semantics, Metamathematics: papers from 1923 to 1938.* Oxford: Oxford University Press, 1956.

[Tarski, 1965] A. Tarski. *Introduction to Logic and the Methodology of Deductive Sciences.* New York: Oxford University Press, 1965.

[Tarski, 1986a] A. Tarski. *Collected Papers*, edited by S.R. Givant and R.N. McKenzie, Basel: Birkhäusen Verlag, 1986a.

[Tarski, 1986b] A. Tarski. What are logical notions? *History and Philosophy of Logic*, 7: 143-154, 1986b (original 1966).

[Tharp, 1975] L. Tharp. Which logic is the right logic? *Synthese*, 31: 1-21, 1975.

[van Benthem, 1982] J. van Benthem. *The Logic of Time*. Dordrecht: D. Reidel, 1982.

[van Benthem, 2001] J. van Benthem. Correspondence Theory. In D. Gabbay and F. Guenthner, (eds.), *Handbook of Philosophical Logic III*, second edition, Dordrecht: Kluwer Academic, 2001.

[van Benthem, forthcoming] J. van Benthem. Introduction. In J. van Benthem *et al.* (eds.), *The Age of Alternative Logics*. Forthcoming.

[van Heijenoort, 1967a] J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book In Mathematical Logic 1879-1931*. Cambridge, Massachusetts: Harvard University Press, 1967a.

[van Heijenoort, 1967b] J. van Heijenoort. Logic as calculus and logic as language. *Synthese*, 17: 324-330, 1967b.

[von Neumann and Morgenstern, 1944] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

[von Wright, 1951a] G.W. von Wright. *An Essay in Modal Logic*. Amsterdam: North-Holland, 1951a.

[von Wright, 1951b] G.W. von Wright. Deontic logic. *Mind*, 60: 1-15, 1951b.

[von Wright, 1983] G.W. von Wright. *Practical Reason: Philosophical Papers*, *Volume 1*. Ithaca: Cornell University Press, 1983.

[Whitehead and Russell, 1912] A.N. Whitehead and B. Russell. *Principia Mathematica*. Cambridge, England: Cambridge University Press, 1912.

[Wittgenstein, 1922] L. Wittgenstein. *Tractatus Logico-Philosophicus*. Translated by C.K. Ogden, London, Routledge and Kegan Paul Ltd, 1922.

[Wittgenstein, 1953] L. Wittgenstein. *Philosophical Investigations*. Oxford: Blackwell, 1953.

[Zeman, 1973] J. Zeman. *Modal Logic*. Oxford: Oxford University Press, 1973.

THE SCOPE AND LIMITS OF LOGIC

Wilfrid Hodges

> The cistern contains, the fountain overflows.
> *William Blake, Proverbs of Hell.*

1 THE BROAD SWEEP

Some features of logic have stayed remarkably constant from Aristotle in the fourth century BC to modern times. We can divide the subject into three broad areas.

> (1) *Theory of inference.* Here we study arguments that are used to justify one statement on the basis of others. Many of these arguments fall into standard argument-patterns, and we classify the patterns.

This is the content of Aristotle's *Prior Analytics*, Ibn Sīnā's *Qiyās*, sections of various Latin medieval texts, the *Dialectica* of Ramus, Boole's *Laws of Thought* and any number of first courses in logic today. Proof theory takes the subject to research level.

> (2) *Theory of definition.* Here we study the use of words or symbols to identify or characterise things; the things can be concrete or abstract. Again there are some standard forms that depend on the kind of thing being characterised.

Definitions were a standard tool of medieval logicians both East and West. We find them discussed by Pascal and Leibniz in the 17th century and by Frege and Peano in the 19th. The subtitle of Barwise's *Admissible Sets and Structures* (1975) is 'An approach to definability theory'. Set theory and model theory revolve around questions of definability. Formal semantics belongs here too.

> (3) *Applied logic.* Here we apply (1) and (2) as tools for analysing or solving practical problems.

For Aristotle the main applications were in scientific explanation and in debate. There are modern counterparts for both of these applications. Proof search (methods for constructing arguments on the fly in order to prove given conclusions) was a major part of logic for Boethius in the 6th century; today it forms one of the main mechanisms of logic programming. Today we also apply logic to teach skills of persuasive writing, and in the design of embedded logic chips.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

The Nyāya logic of India clearly belongs within this general pattern, though there is some debate about whether to place it with (1) as a theory of deduction or with (3) as a theory of scientific explanation. Matilal [Matilal, 2001] presents it as the study of arguments and inference patterns involving the physical location of properties. I know too little about Chinese cultural history to comment on the place of logic in that tradition.

Academic subjects identify themselves partly by social links and sharing of methods. Within the last hundred years statistics has developed its own characteristics and communities, and nobody now places it inside logic. Computability theory has strong links with other parts of mathematical logic and is normally reckoned a part of logic, but this might change if the subject took a swing towards combinatorics. At present cognitive science is set clearly apart from logic by its methods, but some topics of common interest between the two are likely to become more accessible during the next century.

2 DEFINITIONS OF LOGIC

If we want a snappy definition of logic in a few words, it would be hard to improve on Kilwardby's definition from around 1250 [Kilwardby, 1976, para. 523]:

(1) The subject [of logic] is reasoning (*ratiocinatio*).

Kilwardby [Kilwardby, 1976, para. 524] reckoned that this also includes the theory of definition, because reasoning needs definitions.

One might want to add that logic studies good reasoning, not bad. Maybe that goes without saying; piano teachers teach you to play the piano well, not to play it badly. But in any case many logicians have studied bad arguments. Aristotle lectured [Aristotle, 1955] on the skill of producing arguments that are invalid but plausible enough to fool at least some people. Schopenhauer wrote but never published a book [Schopenhauer, 2004] in the same vein. In §5 we will see why modern textbooks of logic no longer have a section on fallacies; the topic is still relevant but it has moved sideways.

Bolzano ([Bolzano, 1837] §7) asks 'Does the question whether coriander improves one's memory belong to logic?' He expects the answer No, and certainly I wouldn't expect to see the question discussed in any logical treatise today. It would be an isolated question with no surrounding theory. Its proper context seems to be the workings of the hippocampus. But logical treatises certainly do discuss equally physiological questions about the parietal lobes, as we will see in §9(b). In the parietal case we have some conscious access to the relevant facts. As our knowledge of the workings of our brains increases, I expect to see closer links between logic and brain science, and I'm not placing bets on how coriander will figure.

There will always be people who denounce parts of logic as 'not real logic', or 'corrupt', or 'charlatan'. To leave my contemporaries at peace I cite Kant, whose *Logik* ([Kant, 1800], written up by Jäsche) starts with a spree of excommunications.

Mathematics is excluded from logic because logic 'may not borrow any principles either from any science or from any experience' [Kant, 1800, p. 529]. Common sense logic 'is not really logic but an anthropological science that has only empirical principles' [Kant, 1800, p. 532]. 'Applied logic really ought not to be called logic' [Kant, 1800, p. 533]. Of the art of producing specious arguments to win a debate, Kant says 'Nothing can be less worthy of a philosopher ... than the cultivation of such an art' [Kant, 1800, p. 531].

Kant's attitude is deplorable. If he wants to single out an area within logic for special consideration, fine; there are many good reasons for doing such a thing. But to reserve the word 'logic' for this selected part is a needless slur on other parts of the subject. There can be shallow or incompetent work done in any area of logic, and most areas allow work of high quality. Much of the rest of this chapter will be a deconstruction of various attempts to narrow down the range of logic.

3 CONCEPTS THAT BELONG TO LOGIC

The phrase 'concepts that belong to logic' appears in Bolzano [Bolzano, 1837, §148]. But a time-honoured way to describe a discipline is to say what its basic concepts are. With logic there is a special problem: we study reasoning, and reasoning is done with concepts. So it's necessary to distinguish between the logician's language with its concepts, and the reasoner's language with its concepts. For convenience let's call the logician's language the *metalanguage* and the reasoner's language the *object language.*

Various logicians at various times have offered lists of concepts belonging to logic. Sometimes these were lists of metalanguage concepts:

(2) ... the logician or dialectician can argue in every special science and in every subject matter, since common concepts, which the dialectician is concerned with, for example genus, species etc., and modes of knowledge and argument, are found in every subject matter.
(The Aristotelian Boethius of Dacia, around 1270, [Boethius of Dacia, 1969, p. 34].)

(3) As the ground of all this procedure is the possibility of our forming the conception of Class the Logic which determines the forms and laws of such procedure, may be said to be the scientific development of the notion of a *class.*
(Boole in about 1860, [Boole, 1997, p. 127].)

Sometimes they were object language concepts:

(4) Just as the concept *point* belongs to geometry, logic likewise has its own concepts and relations ... For example the following belong to logic: negation [i.e. 'not'], identity [i.e. 'equals'], subsumption [i.e. 'is a'], subordination of concepts [i.e. 'every A is a B'].
(Frege in 1906, [Frege, 1906, p. 428].)

There are also medieval lists of words that play roles in arguments.

Debates about the relation between object language and metalanguage concepts go back a long way. In the debate between the Aristotelian Mattā ibn Yūnus and the Islamic scholar As-Sīrāfī before the visier Ibn Al-Furāt in 930 (Elamrani-Jamal [1983, p. 153ff]), one of Sīrāfī's points against Mattā is that a logician has to be prepared to study the concepts that reasoners use in their arguments. He gives as an example the word 'and' (Arabic *wa*); he mentions its roles as a conjunction and as a temporal operator. Mattā is reported as maintaining that Aristotle's metalanguage concepts are adequate for describing all inferences; this was the standard Aristotelian position, as we will see in the next section.

Most object language concepts are unlikely to attract the attention of logicians, because they are not related to any particular argument forms. As Elizabeth Anscombe says [1957, p. 58]:

(5) ... no one has ever pointed out the mince pie syllogism: the peculiarity of this would be that it was about mince pies, and an example would be 'All mince pies have suet in them — this is a mince pie — therefore etc.'

On the other hand, if an object language concept does play a distinctive role in the structure of arguments, then logicians need to be able to name it in their metalanguage too. Those logicians who have tried to prescribe in advance the 'concepts belonging to logic' (either object language or metalanguage) were unwise; who can say what concepts might be found to play a distinctive role in arguments? This doesn't prevent logicians from adopting a particular limited object language for a particular purpose.

One common feature of twentieth century logic has been the use of metalanguage arguments to reach conclusions in the object language. Abraham Robinson's nonstandard analysis is one of the most striking examples of this. Some of the arguments of Hilbert's metamathematics have this form too, though with a different style of metalanguage argument. We will see a further example in §8 below (Gödel's incompleteness theorem). Generally for arguments of this kind it was essential to choose a limited object language and formalise it precisely.

When we discuss the use of metalanguage arguments to reach object language conclusions, we are passing to a meta-metalanguage. This is one illustration of the extraordinary conceptual richness of logic. To try to draw a defining circle around the concepts of logic seems just sad.

4 THE SYLLOGISM CLAIM

Immediately after his statement (1), Kilwardby added that one can also take the subject matter of logic to be 'syllogisms', because [Kilwardby, 1976, para 523]

(6) All modes of reasoning get their force from syllogisms and are reducible to syllogisms.

We can call this the *syllogism claim.* Aristotelian logicians generally accepted it. As a result they took the (metalanguage) concepts of logic to be those needed for describing syllogisms.

Syllogisms are inferences in one of a small number of forms described by Aristotle in his *Prior Analytics* [Aristotle, 1989]. The archetypal syllogistic form is:

(7) If every M is a P and every S is an M, it follows that every S is a P.

Since the middle ages this form (7) has been known as *barbara.* See Thom [1981] for a reliable modern account of syllogisms.

Aristotle himself said [Aristotle, 1989, 68b9–13]:

(8) But now, it should be explained that not only dialectical and demonstrative deductions come about through (*gínontai dia*) the [syllogistic] figures previously mentioned, but also rhetorical ones, and absolutely any form of conviction whatever, arising from whatever discipline.

The passage is a little obscure, like much in Aristotle. But it allows Kilwardby's interpretation, that all inferences are reducible to syllogisms though they may not be syllogisms as they stand.

Maddeningly, neither Kilwardby nor anybody else ancient or medieval tells us in what sense all valid inferences are 'reducible to syllogisms'. If at least they had offered a proof of the syllogistic claim, we could see what they thought they were proving. But for that we have to wait for C. S. Peirce.

In 1867 Peirce offered a proof that 'the general form of every argument' is a syllogism in *barbara* [Peirce, 1867, I §4 with II §3]. He repeated the proof in 1898, in a series of lectures at Harvard [Peirce, 1992, p. 131f]. The following summary is in my words, not Peirce's.

(9) A valid argument is about a subject S, and its conclusion states something about this subject: S is a P. The inference must be warranted by some information about S: S is an M. If M does contain everything about S that is needed to warrant the conclusion, this can only be because of a 'leading principle' relating M and P: Every M is a P.

This reduces the argument to the form

(10) S is an M. Every M is a P. Therefore S is a P.

The form (10) is not exactly *barbara*, but we can bring it to *barbara* by replacing S by the class S' of things that are identical with S. This is a triviality compared with the leaps that Peirce has already made in (9).

Let me make some numbered comments on Peirce's reduction.

(i) Looking back at Leibniz's paper *Elementa Calculi* [Leibniz, 1903, 6] from 1679, we can see in it enough of Peirce's argument to allow the following comments. The more we know about the subject S, the more trivial the clause 'Every M is a

P' can be. The limiting case is where we can define S in enough detail to specify all its properties. For geometry the moral is that we should aim to *define* geometric structures; with adequate definitions, all further mathematical treatment of these structures becomes logic. In support of this picture Leibniz notes that Archimedes eliminated one of Euclid's axioms by defining the relevant notion.

It was shortly after publishing a book on the philosophy of Leibniz that Bertrand Russell gave what we now recognise as a model-theoretic description of geometry [Russell, 1903, ch. xiv]. This was a test case for Russell's claim that 'Pure Mathematics is the class of all propositions of the form "p implies q", where p and q are propositions containing one or more variables, the same in the two propositions, and neither p nor q contains any constants except logical constants.' [Russell, 1903, p. 3]. So even in all its crudity, Peirce's analysis can support real insight into the relationship between mathematics and logic.

(ii) Obviously Peirce's reduction is not the whole truth about logic. But what has he missed? Again Leibniz puts his finger on one essential point [Leibniz, 1962, p. 479]:

(11) It should also be realized that there are valid non-syllogistic inferences which cannot be rigorously demonstrated in any syllogism unless the terms are changed a little, *and this altering of the terms is the non-syllogistic inference.* (My italics)

Exactly so. The same applies in elementary logic when we validate an argument in English by translating it into first-order predicate calculus notation. *The translation itself is a non-first-order inference.* We can formalise more inference steps in first-order logic than we can in syllogisms. But this is a difference of degree, not of kind.

(iii) Peirce himself [Peirce, 1992, p. 131f] points out another important reservation. The problem lies in the 'leading principle'. Before reduction, the original argument inferred that S is a P from the information that S is an M. If this is a sound inference, says Peirce, then there must be a general principle endorsing inferences of this kind, and this general principle is what we write as 'Every M is a P'. But is it clear that we are only entitled to infer from being an M to being a P when every M is a P? Peirce tells us to read the leading principle not in a narrow sense (i.e. literally), but as expressing that some general relation between M and P warrants the passage from premise to conclusion.

No need to spell this out today; default arguments and probabilistic arguments are familiar, not to mention various kinds of scientific inference. Deductive logic is a special case where we can read the leading principle literally.

Comparing himself with Archimedes, Peirce described his analysis of the syllogism claim as the fulcrum 'from which to start a doctrine of inference' [Peirce, 1992, p. 131]. I think it deserves further study.

5 PATTERNS OF REASONING

No academic subject can survive on a diet of isolated facts. Even in history we have to set facts in their proper context and discern the broader picture. Happily the raw material of logic meets us halfway: some patterns of reasoning keep recurring. Take for example the transitivity of 'If . . . then':

(12) If P then Q. If Q then R. *Therefore* if P then R.

At first approximation it seems that when we replace P, Q and R by sentences in this schema, the result must be a cogent argument.

Unfortunately it's not that simple. A famous example goes back to Walter Burley's shorter *De Puritate Logicae* [Burleigh, 1955, p. 203]:

If I say you are a donkey then I say you are an animal.
(13) If I say you are an animal then I say the truth.
Therefore if I say you are a donkey then I say the truth.

So having the syntactic form (12) is not a guarantee of cogency.

The temptation is very strong to say that (12) is obviously a valid argument form, so that in spite of first appearances, Burley's argument (13) is not really of this form. This forces the corollary that valid argument forms are not syntactic but — either in whole or in part — semantic.

Now the syntactic form of an argument is visible for all to see, but the semantic form is not. This opens the door to a line of research in which we identify the semantic requirements for a valid inference. The Latin logicians of the middle ages scoured the examples of bad arguments in Aristotle's *Sophistical Refutations* [Aristotle, 1955], and devised other bad arguments of their own, to see what semantic faults they contained. The outcome was a sophisticated collection of tools for breaking superficially sound arguments. One could claim that the same word stood for different things in different parts of the argument — for example in one place it stood for the objects falling under it, in another place it stood for its meaning and in another it stood for itself. One could claim that quantifiers ranged over different sets in different parts of the argument: here over actual objects, there over possible ones. And so on. (De Rijk [1962].) Burley himself broke the inference (13) by claiming an ambiguity in 'you are an animal'; the words could stand either for a proposition or for themselves [Burleigh, 1955, p. 205].

This could have been a very fertile time for logic, but somehow the energy drained away in the late fourteenth century. Setting aside plagues and politics, there seem to be three reasons internal to logic itself. First, the Latins were extraordinarily naive about speaker's intention and context of utterance. This made it easy to offer several equally plausible refutations of one and the same argument, and even to construct refutations of perfectly good arguments. Renaissance humanists could complain with justice that the scholastics had lost the ability to recognise a valid argument. Thus Thomas More in 1515 [Rogers, 1947, p. 54], after giving pages of examples to support his case:

(14) When heretics are prepared with the same devices as [scholastic theologians], and trained to play the same games, when will they concede? Not quickly, I'll bet, if they show the same contempt for a bunch of twigs as they do for massed arrays of syllogisms.

The moral is that you can't hope to assess the validity of reasoning without some appreciation of the workings of the language in which the reasoning takes place. This is one reason why semantics is an essential tool for logicians.

At this point the second reason for the medieval failure kicked in. Despite the best efforts of the late 13th century modists, the Latin medievals never created a successful theory of syntax. Without a systematic syntax you don't have the basis for a systematic semantics. The writings of Ockham in the early fourteenth century are full of semantic observations, but they don't join up; there is no way you can work through Ockham's text to build up the meanings of complex phrases from those of words.

The third catastrophe followed in turn. While the modists were struggling to explain the syntax of the phrase 'Socrates and Plato' [Covington, 1984, p. 83ff], the nominalists attacked them with a barrage of catch-22 metaphysical questions. A typical example, from Aurifaber around 1330, was 'Every possible thing is either a substance or an accident; so is a mode of signifying a substance or an accident?' [Pinborg, 1967, p. 216]. The science wasn't strong enough to withstand the onslaught. Part of the nominalist case was that all logically significant structure in sentences is at the level of 'mental words'. A thick haze of mental-language logic settled over Europe from the mid fourteenth to the mid nineteenth centuries. Leibniz and a very few other logicians bucked the trend.

Leibniz saw a wholly novel way to deal with problem cases like Burley's (13): invent a language in which they don't happen.

It's by no means obvious that this strategy can work. The problem is the semantics. How can we use English, with its tainted semantics, to define the faultless semantics of the new language?

I believe Kreisel once compared this question with the problem of getting clean dishes by combining dirty dishes and dirty dishwater. Against one's natural expectations, it can be done. The solution was found by generations of mathematicians, honing their English, French or German for precision. One key idea was to take expressions as names of classes wherever possible. Then we need only one primitive relation, namely the relation of class membership; this is the 'subsumption' of Frege's list (4). The first logician to carry out this strategy systematically was Boole in 1847 [Boole, 1847]. For him the English sentence 'No virtuous man is a tyrant' translates into an algebraic sentence saying that X is disjoint from Y, where 'X' and 'Y' are names for the class of virtuous men and the class of tyrants respectively. The reference of 'X' is fixed throughout an argument, and 'X' plays no role except through its reference. Boole noted that the verification of syllogistic arguments, and more besides, becomes routine in the resulting algebra of classes. For example the class-theoretic statement

(15) If $X \subseteq Y$ and $Y \subseteq Z$ then $X \subseteq Z$.

takes the place of *barbara* (6).

In Boole's algebraic language there is no room for the kind of fallacy-chopping we met in the medievals. Simply, there are no fallacies. The schematic sentence (15) becomes true whatever classes we choose for X, Y and Z. (Tarski [1936b] refers to this feature of a sentence schema as 'infallibility' (*niezawodnowość*).) But the medieval problems surface again where Leibniz (11) told us to expect them: in the translation between English and the algebraic language. To see that the translation of an English sentence is correct, we need to understand the semantics of the English sentence. To this problem Boole has contributed nothing at all.

Nevertheless Boole's algebraic language, and the further developments by Jevons, Peirce, Frege, Peano etc., created a revolution in logic. The revolution was mathematical logic. During my own career I remember for example Tony Martin's use of games on the Turing degrees to show ω_1 is measurable under the axiom of determinacy, and Ehud Hrushovski's definition of infinite abelian groups in modular strongly minimal sets. In terms of imagination and sophistication, either of these results beats anything proved in logic before 1900 by many orders of magnitude.

For mathematical arguments the problem of translation into a formal language of logic is often minimal, because modern mathematics is already semi-formal. (Granted, a translation from one formal language to another can still be several years' mathematical research.) In the excitement of mathematical logic one easily forgets the older problems of analysing natural language arguments. But for a variety of reasons, some logicians found themselves revisiting these older problems in the second half of the twentieth century. It would be a tragedy if these researches fell into the same traps that ensnared the medieval efforts. But the situation seems altogether healthier this time round.

First, syntactic theory is a different animal today from the sorry thing it was in the middle ages. Tarski said in 1933 that it would be impossible to handle the semantics of colloquial languages by 'exact methods' without first imposing on these languages 'the characteristic features of the formalized languages' [Tarski, 1983, p. 267]. Today we can take our choice between a range of natural language grammars that largely meet Tarski's requirement.

Second, today's natural language logicians are altogether more realistic about the relevance of context to interpretation. The problems here are still severe, but at least they are being addressed head-on. Computer-based semantic theories, now spreading fast to the web [Fensel *et al.*, 2003], give a kind of reality check: the programs have to work!

Third, today's metaphysicians can still mount a priori demolitions of scientific work, but the research funding bodies and the editors of journals are unlikely to take much notice. This could change in a less friendly political climate. But it will be harder to force this particular genie back into its bottle a second time.

6 REDUCTION TO LOGIC

In a series of works, Frege argued that we can know the truths of arithmetic on the basis of (a) sound definitions, starting from concepts with no empirical content, (b) truths known directly (he called these *Urwahrheiten*, [Frege, 1884, p. 4]), and (c) undeniably correct inference steps. In fact he claimed to demonstrate all this within the (almost) formal language *Begriffsschrift* [Frege, 1879]. This established, he said, that arithmetic needs no proof grounds from either experience or intuition [Frege, 1893, p. 1].

That's quite a mouthful, and of course people abbreviated it. Frege's own abbreviation was that 'Arithmetic is a branch of logic' [Frege, 1893, p. 1]. A later abbreviation described Frege's programme as *logicist*. These slogans had one regrettable consequence, that some discussion of Frege's programme has turned on the proper use of the word 'logic'. If it's just a matter of a name, then why should one care? If there is a real issue behind the label, then it would be better to say what the issue is without using the contentious word. This is why in the previous paragraph I stated Frege's claim without mentioning 'logic'.

Thanks to the Zermelo-Russell paradox, we know today that not all Frege's *Urwahrheiten* were in fact true. So strictly his argument failed. But other logicians proposed similar ways of deriving arithmetic, which broadly seem to work: Whitehead and Russell in *Principia Mathematica*, Zermelo in set theory, Lawvere and Tierney in toposes. For the sake of clarity let me stay with Frege.

In his own words, Frege's aim was epistemological [Frege, 1893, p. vii]:

> (16) The absence of any gaps in the chain of demonstrations has the effect of bringing to light every axiom, presupposition, hypothesis, or whatever you like to call it, that the proof calls on. This gives us a basis for judging the epistemological nature of the law that has been proved.

Of course we can ask the same question 'How do we know?' about any discipline, whether it be microeconomics or cosmology. But in other disciplines we can throw the question back to logic by saying 'We know this by the following reasoning . . . '. In logic there is nowhere else to throw back to. We have the bare reasoning.

Actually Frege makes no contribution (to the best of my knowledge) to the question 'How do we know?' about any particular basic assumption or basic step in reasoning. Instead he sets out the reasoning in a form that he claims is broken down to its atoms. The philosophers of logic can examine the atoms. The fact that his *Urwahrheiten* can't all be true still leaves the question 'How do we know that Axiom V is the wrong one?' So for epistemological purposes his failure is far from total.

It seems to me that Frege's atoms are not at all homogeneous, and they raise different epistemological questions. In a moment I will list a few. Conceivably some philosopher will come to this material with a general explanation of a large class of basic units of reasoning which happens to include all those that are correct in Frege. There might then be some temptation to say that genuine 'logic' is what is covered by this philosopher's explanation. I think we should resist this temptation vigorously. There is no case for the epistemologists of logic to foist their classifications on the rest of us. But in any case I don't believe that any convincing global explanations of this kind are available yet, so the question is premature. I mention it only because some attempts to single out something 'deserving the unqualified name "logic" ' (Feferman [1999, p. 32]) have at least a hint of this motivation behind them.

Frege did in fact try hard to impose some homogeneity on his atoms of reasoning, by the choice of concepts that he made available in *Begriffsschrift*. He criticised Dedekind for relying on 'concepts that are not customary in logic and are not reducible to recognised logical concepts' [Frege, 1893, p. viii], presumably implying that the concepts in *Begriffsschrift* escape this criticism. That looks like the worst sort of appeal to tradition, and it must surprise people who reckon that Frege made fundamental innovations in his choice of logical notions. But I think we can serve Frege better by reading between his lines. He often talks of reduction to logic, but when he is being careful he says that the laws of *Begriffsschrift* are 'pure logical' (*rein logische*, e.g. [Frege, 1879, ix]; [Frege, 1893, p. vii] or 'universal logical' (*allgemein logische*, [Frege, 1884, pp. iv, ix, 4]).

These terms 'pure' and 'universal' hark back to Kant. Kant's explanations seem to me terminally confused (I defy anyone to make sense of [Kant, 1787, p. B84] without special pleading). But Frege himself gives some hints of his meaning. At [Frege, 1879, p. ix] he explains 'pure logical' as meaning 'disregarding the individual features (*besondern Beschaffenheit*) of things'. His logical notions don't quite meet this condition, because he does systematically distinguish between objects and concepts. But the broad aim is clear even if the details are a little untidy: he will use only notions that can properly be applied to any subject matter whatever. (And now we can ask the epistemologist: what is the basis for acquiring such concepts? Or we can ask the brain specialist: how are such concepts built into us, Hurford [2003], Marcus [2001]?)

The least satisfactory part of Frege's scheme is probably his treatment of the *Urwahrheiten*. He seems to assume that if a proposition p uses only subject-free concepts, and we know that p is true, then the ground of our knowledge of p is also subject-free. This is a non sequitur. To illustrate the point, take the second of the basic laws of *Begriffsschrift*, in modern notation:

$$(17)\ \vdash (c \to (b \to a)) \to ((c \to b) \to (c \to a)).$$

How do we know the truth of this law? Answer: by reading the truth-table definition of '→' and then writing out a truth table. This justification uses notions not available in *Begriffsschrift* itself; *Begriffsschrift* has expressions that name truth, but it doesn't express the concept of truth. And even if it did, the proof by truth tables shows that the grounds of our knowledge of the law lie in other laws; it is not an *Ur*-wahrheit.

What has gone wrong here is that we justified (17) by passing to a *different system of representation*, moving up into a metalanguage. Logicians (Gödel [1958]) and cognitive scientists (Stenning [2003]) have rightly emphasised that we can increase our reasoning power by moving from one representation to another. Frege's scheme ignores this issue — or perhaps he hides it from himself by his distinction between how we discover a fact and the grounds of its proof [Frege, 1884, §17]. When we turn to more contentious axioms, for example comprehension axioms or axioms of infinity, then the question 'What concepts are we entitled to use in justifying this axiom?' becomes very pressing.

A similar point applies to Frege's rules of inference.

There are things to say about the justification of Frege's definitions. Frege himself raises a crucial issue in a posthumous paper on 'Logic in Mathematics' [Frege, 1979, p. 208]:

> (18) When something that looks like a definition really makes it possible to prove something which could not be proved before, then it is no mere definition but must conceal something which would have either to be proved as a theorem or accepted as an axiom.

This applies to Frege's definition of 'number'. On one reading, his definition is a declaration that he is going to use the word 'number' as an abbreviation for a longer expression of *Begriffsschrift*. In that case the definition is redundant, and nothing has been said about the concept of number. This is clearly not what Frege wanted. The alternative is that the definition is a true statement expressing a logical analysis of the concept of number. In that case, says Frege [Frege, 1979, p. 209f.],

> (19) We cannot prove [that the analysis is correct]. The most one can be certain of is that as far as the form of words goes we have the same sentence after the analysis as before. But that the thought itself also remains the same is problematic. . . . I believe that we shall only be able to assert that [this is so] when this is self-evident. And then what we have is an axiom.

In short, the epistemological question arises for Frege's definitions just as much as for his logical axioms, though in a different form.

Both the question of shift of representation in proofs and the question of the self-evidence of logical analyses seem to me enormously perplexing. I hope that closer interactions between logicians and cognitive scientists will throw light on both of them. Rigour in the formalisation will be essential.

7 QUINE'S 'SCOPE OF LOGIC'

'The scope of logic' is the title of Chapter 5 of Quine's *Philosophy of Logic* [Quine, 1970].

Quine's chapter has a context. Earlier in the book Quine has shown how to set up a formal language with a set-theoretic semantics. He has explained how to define the notions of logical implication and logical incompatibility in terms of the notion of logical truth, and he has suggested a way of defining logical truth in terms of the grammar and semantics of the formal language: 'A *logical truth* is a *sentence that cannot be turned false by substituting for lexicon*' ([Quine, 1970, p/ 58], his italics). It remains to pin down which expressions of the language are lexical. He concentrates on predicates, and he asks which predicates are 'proper to logic' and which are 'extra-logical' (i.e. lexical).

Quine is picking out a certain area of enquiry within logic as a whole. Unless his selection happens to be interesting for some accidental reason, the first sensible question about it is whether it meets the purpose he had in making it. Sadly he never tells us that purpose, so we have to infer it from the arguments that he uses.

What counts in favour of a predicate P being proper to logic?

(a) Counting P as proper to logic doesn't prevent our having a complete proof calculus for the logical truths [Quine, 1970, pp. 62, 72, 82f].

(b) Counting P as proper to logic doesn't commit us to logical truths that express 'staggering existential assumptions' [Quine, 1970, p. 68]. In the same vein, counting P as proper to logic doesn't give us logical truths 'expressible by direct quantification in the object language' [Quine, 1970, p. 61].

(c) Counting P as proper to logic doesn't prevent a consensus on what sentences are logical truths [Quine, 1970, p. 72] 'competing doctrines').

(d) The predicate P 'treats of all objects impartially' [Quine, 1970, p. 62].

In his next chapter Quine describes the disadvantages of using a 'deviant logic'. He argues against deviant logics on the grounds that they lack 'simplicity' and are less 'familiar' [Quine, 1970, p. 86]. These criteria could also be relevant to choosing what predicates are proper to logic.

Item (d) is familiar from §6 above. But where do the other criteria come from? When I first read Quine I was bewildered by them. I think I suggested in print somewhere that the cash value of Quine's question was 'What formal system of logic should we teach to undergraduate logic classes?' Only later I came across Bolzano's definition of logic as 'the science that shows us how to present the sciences in adequate textbooks (*Lehrbüchern*)' [Bolzano, 1837, §1].

Knowing more of the history of logic than I used to, I think I can put Quine's question into a context where his criteria make sense, but I can't guarantee this is the only way of solving the jigsaw puzzle.

Around the beginning of the twentieth century, various trends converged into a consensus about the proper methodology for deductive disciplines. For the sake of a name I call this consensus the *logistic paradigm.* Each discipline should have its own primitive concepts and its own axioms; these should be expressible in some suitable formal language. The discipline should be developed by adding concepts through formal definitions and adding theorems through formal derivations. In this form the discipline was called a *deductive theory.* One can find descriptions of this methodology in Hilbert [1918], Weyl [1928], Tarski [1936a] and Church [1956, §07].

What should the formal proof calculus be? Tarski in [1933b] left this open; each formal theory should come equipped with its own logical axioms and inference rules, and these need not be the same in all systems. For one particular formal theory Hilbert [1931] proposed a proof rule with infinitely many premises. In another direction, several authors stressed the advantages of having a notion of derivation which is effective, in the sense at least that the set of deducible theorems is effectively enumerable. For example Church [Church, 1956, p. 52] argued that 'to the extent that requirements of effectiveness fail, the purpose of communication is defeated', because the formal theory fails to communicate its theorems to its users. A more technical argument with the same conclusion was that Hilbert's metamathematics could only be applied to theories with a proof system of a standard type.

There would be obvious advantages if all deductive scientists could agree to use the same logic, so that the differences between one deductive theory and another lay in the primitive concepts and the non-logical axioms. It would also be helpful if the logical consensus included only things that all theories have to assume, so that special claims from one discipline don't creep into others by stealth.

In the context of the logistic paradigm, all of Quine's criteria make perfect sense. Quine supposes that we are in the business of setting up formal systems to cope with any branch of knowledge that we want to formalise. The logic should be kept small and uncontroversial, and the set of logical truths should be computably enumerable. Simplicity, familiarity and avoidance of controversial issues are all merits in a logic that the community is going to agree to use for all sciences.

The snag is, by 1970 everybody except Quine had abandoned the logistic paradigm. I don't think it collapsed under one particular assault, like the walls of Jericho. But Stegmüller points out the impracticality of the paradigm [Stegmüller, 1979, p. 5]:

> (20) But suppose that Bourbaki had made up his mind to use, as a universal language of mathematics, *formalized set theory* instead of *informal set theory* for his intended reconstruction of modern mathematics in precise terms. The Bourbaki volumes would then look like, for example, the last chapter of Shoenfield's *Mathematical Logic.* I really fear that the Bourbaki volumes would not exist, at least not yet, since instead of having published more than 20 volumes, Bourbaki would still be working on his first volume.

Informal first-order Zermelo-Fraenkel set theory survives as the agreed foundations for most mathematicians who care about foundations. (I wrote this on 5 March 2005.) Where one needs a formal theory for a particular purpose — for example for some Hilbert-style metamathematics like quantifier elimination — one can do the job within the universe of sets; but most of Quine's criteria will be irrelevant to the choice of logic for the theory.

And of course there are the textbooks of logic. Quine's criteria of familiarity and simplicity have a lot going for them in undergraduate classes. The choice of classical first-order logic as the logic to teach students owes as much to Gerhard Gentzen's limpid proof methods as to anything else.

In short, Quine's question about the scope of logic is dead in the water. But before we leave the logistic paradigm, we should look at some things that Tarski said in the same connection.

8 LOGICAL NOTIONS

In 1936 Tarski published a paper [Tarski, 1936b] that proposed a way of defining 'logical consequence' in the context of deductive theories (i.e. the logistic paradigm).

Tarski's paper distinguishes at various points between three notions of logical consequence. Notion One (my name) is 'colloquial' (*potoczny*) or 'common' (*üblich*), and Notion Two is the formalisation of this colloquial notion within the logistic paradigm. (He draws this contrast also in the last two pages of [Tarski, 1933a].) Notion Three is his own proposal for making Notion One precise. Today Notion Two is recognisable as proof-theoretic consequence; strictly it may differ from one deductive theory to another, but there was some consensus at least about the rules for second-order logic. Notion Three is today's semantic consequence. Thanks largely to Tarski's own work, this notion is now just as formalised as the proof-theoretic one.

Tarski claimed that Notion Two failed to exhaust the content of Notion One. His most telling piece of evidence for this was Gödel's incompleteness theorem (though his account of it is rather laconic). Consider pure second-order logic $\mathcal{L}$, presented as a formal system. In this system we can define the natural numbers and prove all the axioms of first-order Peano arithmetic. So by Gödel's results, assuming the truth of the axioms of $\mathcal{L}$, we can use a metalanguage argument to establish the truth of a statement in the language of $\mathcal{L}$ that is not formally provable in $\mathcal{L}$. This shows an ambiguity in the notion of a 'provable statement of $\mathcal{L}$'; in a narrow sense, it means a proof-theoretic consequence of the axioms of $\mathcal{L}$, whereas in a stronger sense it includes statements of $\mathcal{L}$ proved in the metatheory. This stronger sense is open-ended; other arguments that nobody has yet thought of might establish still further truths in $\mathcal{L}$. In Tarski's view, the stronger sense is included in Notion One; so Notion One is broader than Notion Two, and it needs to be made precise independently from Notion Two.

Tarski took as his starting point the notion of 'infallible' schemas that we met in §5 above. Suppose K is a class of sentence schemas and X is a sentence schema, and suppose that every possible assignment of references to the schematic letters making all of K true also makes X true. Today we express this by saying that X is a semantic consequence of K. A universally agreed requirement on a proof rule for a deductive theory was that its conclusion must be a semantic consequence of its premises. With one important rider, Tarski's proposal in [Tarski, 1936b] was to shift the name 'logical consequence' from proof-theoretic consequence to semantic consequence.

The rider was that in deductive theories the sentences are not schemas; they are fully meaningful sentences. Tarski's proposal was that a deductive theory should include a distinction between those of its expressions or concepts that are 'logical' and those that are not. (He already followed this practice in his own work of this period.) Then a sentence X should be regarded as a 'logical consequence' (in his revised sense) of the set of sentences K if replacing all the non-logical expressions of K and X systematically by variables turns them into a set of schemas K' and a schema X' such that X' is a semantic consequence of K'.

This rider dropped out of use around 1950, when model theory (mostly using first-order languages with built-in schematic letters) replaced the logistic paradigm. For first-order languages proof-theoretic consequence coincides with semantic consequence, thanks to the completeness theorem. There are formal theories in various application areas, with their own languages and logics. For example there are formal theories of processes; these theories are in languages with expressions that stand for general features of processes (e.g. an operator on sentence ϕ, expressing that an action of a certain type takes the present state to one in which ϕ holds). These expressions are declared to be 'logical', and the result is 'process logic'. This approach vindicates Tarski's decision to leave it open which concepts are 'logical', though probably not in a way that he foresaw in 1936.

Tarski described Notion One as 'colloquial', 'common usage' and as having an 'intuitive content'. Some readers of [Tarski, 1936b], fifty years later, inferred that he was talking about a common intuition and offering his own conceptual analysis of this intuition. If that were true, it would be highly relevant to the topic of this chapter. We could all introspect our common intuition of logical consequence and compare our analyses of it with Tarski's.

But of course there is not and never has been a common intuition of logical consequence, any more than there is a common intuition of inductive definition or a common intuition of subatomic physics. To me it's extraordinary that readers who thought Tarski meant there is such an intuition never challenged him on this. Where was Tarski's evidence of the existence of this intuition? Whatever did they think they were introspecting?

Tarski's meaning is not entirely clear. In places he speaks of Notion Two as 'new' in contrast with Notion One, or as 'old' relative to Notion Three, suggesting that the Notion One he has in mind is older than the formal notions of consequence within deductive theories. On the other hand his invocation of syntactic form suggests a fairly recent notion. My guess is that he meant the notion (whatever it was) that the founders of the logistic paradigm had in mind when they formulated their proof rules. This notion may only have been 'common usage' within the community of logicians, and it was never 'colloquial' except in the sense that as an informal notion it had to be expressed in colloquial language.

Tarski doesn't cite his evidence for Notion One. The historical record seems thin — the phrase 'logical consequence' hasn't much of a pedigree. When Frege talked of a 'logische Folge' in the introduction to *Begriffsschrift* [Frege, 1879], he was referring to the sequence of natural numbers. Weyl's [1928] notion of a 'logische

Folge' of a set of axioms seems to be a statement derivable from the axioms by normal mathematical methods of deduction. Hilbert and Ackermann [1928, p. 1] in 1928 speak of a sentence in a deductive theory being a 'logical consequence' (*logische Folgerung*) of other sentences if it's derivable from the other sentences by formal proof rules of the deductive theory. This is already Notion Two. Neither Weyl nor Hilbert and Ackermann provide much support for Tarski's claim to be explicating Notion One.

By 1967, when Tarski proposed a definition of 'logical' notions, he expressed himself more cautiously [Tarski, 1986, p. 145]:

> (21) Let me tell you in advance that in answering the question 'What are logical notions?' what I shall do is make a suggestion or proposal about a possible use of the term 'logical notion'. This suggestion seems to me to be in agreement, if not with all prevailing usage of the term 'logical notion', at least with one usage which actually is encountered in practice. I think the term is used in several different senses and that my suggestion gives an account of one of them.

I would rather Tarski had chosen a name less likely to suggest that he was solving a substantive philosophical problem. But maybe his earlier paper [Tarski, 1936b] forced this name on him.

9 HOW WE OUGHT TO THINK?

Kant said [1800, p. 529 of the translation]:

> (22) In logic, however, the question is not about *contingent* but about *necessary* rules; not how we do think, but how we ought to think.

Unless Kant is being unpleasantly moralistic, his point here must be that a knowledge of logic equips us to think in ways that are necessary for achieving some goal that we all share. Compare the nutritionist. Her expertise qualifies her to say that if we want to stay healthy, we ought to eat this and avoid that.

What is to thinking as health is to eating? I have no idea. There are so many things we can achieve by thinking: finding our way around, avoiding predators, impressing members of the opposite sex. The question is complicated by the fact that we can build machines to think for us, for any purpose we choose. In view of all this I can't offer any very systematic discussion of the question, but let me make a few points.

(a) We can start at the other end and ask what we can do better as a result of a training in logic. Luce in the introduction to his Aristotelian *Teach Yourself Logic* [Luce, 1958, p. vii] says:

> (23) In lecturing on Logic to University men and women, over and over again I have seen the same door swing open, the same step forward taken, the same marked development. The nerveless, juvenile letter written to the College Tutor twelve months previously, has now turned into a logical, well-knit communication, its subject-matter thought out, its phrasing clear and concise. Logic can do this for you.

I'm no more qualified to pronounce on these things than Luce was, but I'm highly sceptical of his last sentence.

True, a training in syllogisms might well be better for learning transferable reasoning skills than a course in first-order logic. In both cases the particular skill of manipulating the formalism is extremely specialised and probably doesn't transfer to anything else at all. The general educational benefit probably comes from the part that Leibniz (11) emphasised, the translation between English arguments and the formal system. Since the syllogistic forms are cruder than those of first-order logic, students of syllogisms will spend a larger proportion of their time on the useful translations and less on the useless symbol manipulation. This seems to accord with the thinking of 'informal logicians' who use logic as a framework for teaching communication and discussion skills. They tend to use only simple argument forms (not necessarily syllogisms), and their 'logical' concepts are almost all at metalanguage level. (See for example Groarke [].)

I hasten to add that for mathematics and computer science students, the formal manipulations of first-order logic are much closer to their daily business, and so the formal skill is closer to the applications. I teach a second-year undergraduate course [Hodges, 2004] that uses logic as a framework for mathematical writing; the course is built around an informal use of the rules of natural deduction together with some items in the theory of definition.

(b) The safest assumption might be that the goal Kant has in mind is close to home, and he means that logic can teach us how to think if we are going to *think well.*

Two marks of thinking well are *accuracy* and *speed.* Both requirements come to the fore when one is designing safety-critical devices, for example programmable logic devices for a critical system. The UK Ministry of Defence used to require safety-critical systems to be validated by formal methods of logic; this caused problems because the formal methods were too slow unless one took short cuts that threatened their reliability. The latest UK standard [Def, 2004] has a different emphasis. The contractor is now required (9.1f) to produce a 'structured argument, supported by a body of evidence' to show that the system is safe for the intended use; nothing is said about the form of the argument. The contractor is also required to estimate risks, and to keep a log of failures and near-misses.

If we move from chips that think to humans who think, then we are talking about the behaviour of the cerebral hemispheres. I confine myself to the question of speed here. One reason that Euler diagrams appear in some logic texts is that they allow extremely fast visual processing. Almost certainly the explanation lies in the parietal lobes. Logicians also discuss how to make symbolic calculations fast and efficient on the page; for example the 'fell swoop' in Quine's *Methods of Logic* ([Quine, 1952, p. 35], noting his remark about what is 'visibly verifiable').

(c) Under (b) I talked mainly about reasoning of a technical kind. Some people have argued that there is no need to teach logic of the kind that we use in everyday inferences, because we never make mistakes of logic in this kind of reasoning. Thus Descartes [Descartes, 1966, §2]:

> (24) [For pure deductions] I see little to be gained from those shackles by which the Logicians reckon to control human reason, though I wouldn't deny that they are highly suitable for other applications. In fact all the errors that people are subject to (I'm not speaking of dumb animals) are in every case the result not of faulty inference, but of inadequate understanding of the observed facts, or of hasty and unjustified judgement.

It has turned out to be surprisingly hard to establish what if anything goes wrong when people are invited to reason and seem to do something different. Did they misunderstand the instructions, or the sentences, or were they not trying to do what they were asked, or what? To sort out these questions needs a delicate appreciation both of logical principles and of the realities of an experimental situation. I commend Stanovich [1999]; his conclusion is that people do make mistakes of propositional logic ('there were reliable deviations from normative responding', [Stanovich, 1999, p. 208], but he is right to go cautiously.

Where Descartes sees 'hasty' judgement, Stanovich points to 'computational limits'. These are just as relevant to everyday reasoning as they are to safety-critical software. It's unhelpful to say we 'ought' to think in a certain way if that way of thinking is beyond our capacity in the time available.

A few years ago an American cognitive scientist invited me to join her in a project on the use of working memory space in arguments from assumptions. It would have involved MRI scans of people who find it difficult to carry out such arguments. I very much regret not having had the time to take up this offer.

(d) The study of how people actually do reason helps to nail some misdescriptions of logic. It used to be common to say that the chief aim of logic is to bring us to true beliefs. But consider this report of an exchange between an Experimenter from Moscow State University and an illiterate Kashgar peasant called Abdurahman (Luria [1976, p. 108f.]; I compressed it a little):

> **E.** In the Far North, where there is snow, all bears are white. Novaya Zemlya is in the Far North and there is always snow there. What color are the bears there?
>
> **A.** We always speak only of what we see; we don't talk about what we haven't seen.
>
> **E.** But what do my words imply?
>
> **A.** Well, it's like this: our tsar isn't like yours, and yours isn't like ours. Your words can be answered only by someone who was there, and if a person wasn't there he can't say anything on the basis of your words.

You can make your own analysis of this dialogue. But it's clear which of the participants is aiming at truth, and it's not the Experimenter. The Experimenter is aiming at valid deduction. Obviously there are connections between reaching truth and deducing validly, but they are indirect and easily misunderstood.

(e) One often reads that Frege and Husserl eliminated psychology from logic through their attacks on psychologism. This description obscures the facts.

Frege attacked the use of psychology in logic throughout his *Grundlagen* [Frege, 1884] and in other places. Generally his main target is bad logic. The *Grundlagen* is about how to define 'number'. Frege rightly points out that before we tackle questions of content, a definition of 'number' must at least have an appropriate format. That format should show how the concept 'number' ties in with other concepts in propositions that use it. So at least it must show the (linguistic) argument structure of the concept. He is clearly correct that we can't read off the argument structure from a mental picture. [Frege, 1884, §§45–48, 58–60]. But against Frege, this is a fact about pictures, not about minds. A children's dictionary that defines words by pictures won't tell us the argument structures any better than a mental image. And conversely we do have the concept 'number' in our minds, argument structure and all; there is more to our minds than mental pictures.

In short, there is no reason of principle why a competent psychological investigator of the meanings of words should disagree with Frege's analysis of the argument structure of 'number'. Frege's target was a type of mentalistic logician, and it has to be said that the track record of mentalistic logicians from Peter of Ailly to Sigwart is unimpressive. One of the worst features of their work was their inappropriate appeal to introspection. I mentioned in §8 above that some philosophical logicians today make equally unacceptable appeals to introspection, though the issues in their case are not those discussed by Frege.

On the other hand Gödel had no hesitation in appealing to introspection as a source of concepts and evidence in logic. Nobody is going to describe Gödel's track record in logic as poor. So perhaps the most significant question raised by Frege's critique is a methodological question about the proper use of introspection.

Husserl came at this question more directly than Frege did. For Husserl the subject matter of logic consists of some things that we do with our minds [Husserl, 1900, §18]:

> (25) Concepts, judgements, syllogisms, deductions, inductions, definitions, classifications etc. — all psychology, except that they are selected and arranged from normative and practical points of view.

His problem is how to use our introspective evidence of these operations of our own minds, as grounds for knowledge of 'ideal, necessary connections' (e.g. [Husserl, 1900, §56]. Things have moved on since Husserl's time. For example Chomsky insists that language is a mental phenomenon, and he has constantly used introspective evidence to support his linguistic theories. In his view, false empiricist methodologies have done more harm than use of any particular source of evidence. (Chomsky [1986], particularly Ch. 2 'Concepts of Language'.)

As a working logician I'm happy to leave these foundational questions to deeper minds than mine.

BIBLIOGRAPHY

[Anscombe, 1957] G. E. M. Anscombe. *Intention.* Blackwell, Oxford, 1957.

[Aristotle, 1955] Aristotle. *Sophistical Refutations.* Trans. E. S. Forster. Loeb Classical Library. Harvard University Press, Cambridge Mass., 1955.

[Aristotle, 1989] Aristotle. *Prior Analytics, trans. Robin Smith.* Hackett, Indianapolis IN, 1989.

[Bolzano, 1837] B. Bolzano. *Wissenschaftslehre: Versuch einer ausführlichen und grösstentheils neuen Darstellung der Logik mit steter Rücksicht auf deren bisherige Bearbeiter.* Seidel, Sulzbach, 1837. Partial translation as *Theory of Science*, ed. Rolf George, University of California Press, Berkeley 1972.

[Boole, 1847] G. Boole. *The Mathematical Analysis of Logic, being an Essay towards a Calculus of Deductive Reasoning.* Macmillan, Barclay and Macmillan, Cambridge, 1847.

[Boole, 1997] G. Boole. Logic. In Ivor Grattan-Guinness and Gérard Bornet, editors, *George Boole: Selected Manuscripts on Logic and its Philosophy*, pages 126–153. Birkhäuser, Basel, 1997.

[Boethius of Dacia, 1969] Boethius of Dacia. Modi Significandi sive Quaestiones Super Priscianum Maiorem. In Joannes Pinborg and Henricus Roos, editors, *Boethii Daci Opera, vol. iv part 1.* Gad, Copenhagen, 1969.

[Burleigh, 1955] W. Burleigh. *De Puritate Artis Logicae, Tractatus Longior.* Ed. Philotheus Boehner. The Franciscan Institute, St. Bonaventure, New York, 1955.

[Chomsky, 1986] N. Chomsky. *Knowledge of Language, its Nature, Origin, and Use.* Praeger, New York, 1986.

[Church, 1956] A. Church. *Introduction to Mathematical Logic vol. I.* Princeton University Press, Princeton, 1956.

[Covington, 1984] M. A. Covington. *Syntactic Theory in the High Middle Ages: Modistic Models of Sentence Structure.* Cambridge University Press, Cambridge, 1984.

[de Rijk, 1962] L. M. de Rijk. *Logica Modernorum i: On the Twelfth Century Theories of Fallacy.* Van Gorcum, Assen, 1962.

[Def, 2004] Defence Procurement Agency, Glasgow. *Safety Management Requirements for Defence Systems Part 1: Requirements*, 2004.

[Descartes, 1966] R. Descartes. *Regulae ad Directionem Ingenii, ed. G. Crapulli.* Nijhoff, The Hague, 1966.

[Elamrani-Jamal, 1983] A. Elamrani-Jamal. *Logique Aristotélicienne et Grammaire Arabe.* J. Vrin, Paris, 1983.

[Ewald, 1996] W. Ewald. *From Kant to Hilbert, Vol. II.* Clarendon Press, Oxford, 1996.

[Feferman, 1999] S. Feferman. Logic, logics, and logicism. *Notre Dame Journal of Formal Logic*, 40:31–55, 1999.

[Fensel *et al.*, 2003] D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster. *Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential.* MIT Press, Cambridge Mass., 2003.

[Frege, 1879] G. Frege. *Begriffsschrift: Eine der Arithmetischen Nachgebildete Formelsprache des Reinen Denkens.* Nebert, Halle, 1879.

[Frege, 1884] G. Frege. *Die Grundlagen der Arithmetik.* Koebner, Breslau, 1884.

[Frege, 1893] Gottlob Frege. *Grundgesetze der Arithmetik Begriffsschriftliche Abgeleitet, I.* Pohle, Jena, 1893.

[Frege, 1906] G. Frege. Über die Grundlagen der Geometrie III. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 15:423–430, 1906.

[Frege, 1979] G. Frege. Logic in mathematics. In Hans Hermes, Friedrich Kambartel, and Friedrich Kaulbach, editors, *Gottlob Frege, Posthumous Writings*, pages 203–250. Blackwell, Oxford, 1979. Translation of unpublished notes 'Logik in der Mathematik', 1914.

[Gödel, 1958] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958. Reprinted with translation in *Kurt Gödel: Collected Works II, Publications 1938–1974*, ed. Solomon Feferman et al., Oxford University Press, New York 1990, pp. 240–251.

[Groarke,] L. Groarke. Informal logic. Stanford Encyclopedia of Philosophy, `plato.stanford.de` as of 9 March 2005.

[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der Theoretischen Logik.* Springer Verlag, Berlin, 1928.

[Hilbert, 1918] D. Hilbert. Axiomatisches Denken. *Mathematische Annalen*, 78:405–415, 1918. Translated as 'Axiomatic Thought' in [Ewald, 1996] pp. 1107–1115.

[Hilbert, 1931] D. Hilbert. Die Grundlegung der Elementaren Zahlenlehre. *Mathematische Annalen*, 104:485–494, 1931. Translated as 'The grounding of elementary number theory' in [Ewald, 1996] pp. 1148–1157.

[Hodges, 2004] W. Hodges. An experimental course in mathematical writing. *MSOR Connections*, 4:29–31, 2004.

[Hurford, 2003] J. R. Hurford. The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, 26:261–316, 2003.

[Husserl, 1900] E. Husserl. *Logische Untersuchungen I: Prolegomena zur Reinen Logik.* Max Niemeyer Verlag, Tübingen, 1900.

[Kant, 1787] I. Kant. *Kritik der Reinen Vernunft.* Hartknoch, Riga, 1787.

[Kant, 1800] I. Kant. *Logik: Ein Handbuch zu Vorlesungen herausgegeben von Gottlob Benjamin Jäsche.* Friedrich Nicolovius, Königsberg, 1800. Trans. as *The Jäsche Logic* in Immanuel Kant, *Lectures on Logic*, ed. J. Michael Young, Cambridge University Press, Cambridge 1992.

[Kilwardby, 1976] R. Kilwardby. *De Ortu Scientiarum, ed. Albert G. Judy.* British Academy, London, 1976.

[Leibniz, 1903] G. W. Leibniz. Elementa Calculi. In Louis Couturat, editor, *Opuscules et Fragments Inédits de Leibniz*, pages 49–57. Alcan, Paris, 1903.

[Leibniz, 1962] G. W. Leibniz. *Nouveaux Essais sur l'Entendement Humain, ed. André Robinet and Heinrich Schepers.* Akademie-Verlag, Berlin, 1962.

[Luce, 1958] A. A. Luce. *Logic.* Teach Yourself Books. English Universities Press, London, 1958.

[Luria, 1976] A. R. Luria. *Cognitive Development: Its Cultural and Social Foundations.* Harvard University Press, Cambridge Mass., 1976.

[Marcus, 2001] G. F. Marcus. *The Algebraic Mind: Integrating Connectionism and Cognitive Science.* MIT Press, Cambridge Mass., 2001.

[Matilal, 2001] B. K. Matilal. Introducing Indian logic. In Jonardon Ganeri, editor, *Indian Logic: a Reader*, pages 183–215. Curzon, Richmond, Surrey, 2001.

[Peirce, 1867] C. S. Peirce. On the natural classification of arguments. *Proceedings of the American Academy of Arts and Sciences*, 7:261–287, 1867. Reprinted in *Writings of Charles S. Peirce, Vol. 2, 1867–1871*, ed. Edward C. Moore et al., Indiana University Press, Bloomington IN 1982, pp. 23–48.

[Peirce, 1992] C. S. Peirce. *Reasoning and the Logic of Things, ed. Kenneth Laine Kettner.* Harvard University Press, Cambridge Mass., 1992.

[Pinborg, 1967] J. Pinborg. *Die Entwicklung der Sprachtheorie im Mittelalter.* Aschendorffsche Verlagsbuchhandlung, Munster, 1967.

[Quine, 1952] W. V. O. Quine. *Methods of Logic.* Routledge and Kegan Paul, London, 1952.

[Quine, 1970] Willard Van Orman Quine. *Philosophy of Logic.* Prentice-Hall, Englewood Cliffs NJ, 1970.

[Rogers, 1947] E. F. Rogers. *The Correspondence of Sir Thomas More.* Princeton University Press, Princeton NJ, 1947.

[Russell, 1903] B. Russell. *The Principles of Mathematics.* George Allen and Unwin, London, 1903.

[Schopenhauer, 2004] A. Schopenhauer. *The Art of Always Being Right: Thirty Eight Ways to Win When You are Defeated.* Gibson Square Books, London, 2004.

[Stanovich, 1999] K. E. Stanovich. *Who is Rational?* Lawrence Erlbaum, Mahwah NJ, 1999.

[Stegmüller, 1979] W. Stegmüller. *The Structuralist View of Theories.* Springer-Verlag, Berlin, 1979.

[Stenning, 2003] K. Stenning. *Seeing Reason.* Oxford University Press, Oxford, 2003.

[Tarski, 1933a] A. Tarski. Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit. *Monatshefte für Mathematik und Physik*, 40:97–112, 1933. Reprinted in *Alfred Tarski, Collected Papers vol. 1 1921–1934*, ed. Steven R. Givant and Ralph N. McKenzie, Birkhäuser, Basel 1986, pp. 621–636.

[Tarski, 1933b] A. Tarski. *Pojęcie Prawdy w Językach Nauk Dedukcyjnych*, volume 34 of *Prace Towarzystwa Naukowego Warszawskiego.* Wydział III Nauk Matematyczno-Fizycznych, Warsaw, 1933. Reprinted [Tarski, 1995, pp. 13–172]; translated as 'The concept of truth in formalized languages' in [Tarski, 1983] pp. 152–278.

[Tarski, 1936a] A. Tarski. *O Logice Matematycznej i Metodzie Dedukcyjnej*, volume 3–5 of *Biblioteczka Mat.* Książnica-Atlas, Lwów and Warsaw, 1936. Translated as *Introduction to Logic and to the Methodology of the Deductive Sciences*, translation and revision of [Tarski, 1936a], Oxford University Press, New York 1941.

[Tarski, 1936b] A. Tarski. O pojęciu wynikania logicznego. *Przegląd Filozoficzny*, 39:58–68, 1936. Reprinted [Tarski, 1995; pp. 186–202]; translated by Tarski as 'Über den Begriff der logischen Folgerung', in *Actes du Congrès International de Philosophie Scientifique* 7, Herman et Cie, Paris 1936, pp. 1–11; translated as 'On the concept of logical consequence' in [Tarski, 1963b, pp. 409–420].

[Tarski, 1983] A. Tarski. *Logic, Semantics, Metamathematics: papers from 1923 to 1938, ed. John Corcoran.* Hackett Publishing Company, Indianapolis IN, 1983.

[Tarski, 1986] A. Tarski. What are logical notions? *History and Philosophy of Logic*, 7:143–154, 1986.

[Tarski, 1995] A. Tarski. *Pisma Logiczno-Filozoficzne, Tom 1 'Prawda', ed. Jan Zygmunt.* Wydawnictwo Naukowe PWN, Warsaw, 1995.

[Thom, 1981] P. Thom. *The Syllogism.* Philosophia, Munich, 1981.

[Weyl, 1928] H. Weyl. *Philosophie der Mathematik und Naturwissenschaft.* Leibniz Verlag, Munich, 1928.

LOGIC IN PHILOSOPHY

Johan van Benthem

1 THE CENTURY THAT WAS

Logic has played an important role in modern philosophy, especially, in alliances with philosophical schools such as the Vienna Circle, neo-positivism, or formal language variants of analytical philosophy. The original impact was via the work of Frege, Russell, and other pioneers, backed up by the prestige of research into the foundations of mathematics, which was fast bringing to light those amazing insights that still impress us today. The Golden Age of the 1930s deeply affected philosophy, and heartened the minority of philosophers with a formal-analytical bent. As Brand Blanshard writes in *Reason and Analysis* [1964] — I quote from memory here, to avoid the usual disappointment when re-reading an original text:

> *"It was as if a little band of stragglers, in disarray after a lost battle, suddenly found Napoleon's legions marching right alongside of them..."*

In the 1940s and 1950s, people like Carnap, Reichenbach, Quine, and their students made logic-based methodology a highly visible *modus operandi* in the field. Then, in the 1960s, what came to be called 'philosophical logic' began to flourish, and logicians like Hintikka, Geach, Dummett, Kripke, Lewis, and Stalnaker came to prominence, not just in an auxiliary role as logical lawyers (Prior's phrase), but even setting the agenda for philosophical discussions that were now suddenly considered mainstream.

It seems fair to say that this age of logical influence has waned. Many philosophical circles have no live contact with logic any more, and basic logical skills seem scarce in the current canon of 'what every philosopher should know'. Even technically inclined philosophers often feel that probability or game theory or dynamical systems are more versatile tools than logic for making formal points and relevant distinctions. And for their part, many logicians have gone elsewhere, to areas with more lively interactions, in departments of computer science, linguistics, or even further astray in the social sciences. Indeed, the bulk of logic research today is done in computer science, as can easily be checked in terms of numbers of publications, journals, organizations, and sheer variety of research topics.

Reactions to this situation vary. One is a turn to writing the history of the glory days of the foundational era, often with the concomitant fundamentalist view that everything would be better nowadays if we just returned to the purity

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

— largely mythical — of that lost golden age. A response with which I feel much more sympathy is the activist stance of reintroducing logical ideas into philosophy — contemporary ones, to be sure. Indeed, the gap between current philosophical practice and modern logic may be more apparent than real. Many themes still run on parallel lines, and some differences just arise for practical reasons, viz. the absence of textbooks explaining modern logic to a philosophical audience, or of influential publications in philosophy showing modern logic at work. But take any informal gathering, and it soon turns out that philosophers and logicians are still on speaking terms, and sometimes even more than that!

In this chapter, we do not take a particular stand, but just try to approach the whole issue in a way that is somewhat different from the usual approaches.

Four dead-end streets

In setting out this role of logic in philosophy, however, there are some situations to be avoided that obstruct an impartial view.

One are broad *philosophized histories* of what took place. A good example is the recent popularity of schematic views of the history of logic which claim that the field has been heading to a 'meta-level' self-understanding in the 1960s, occupied mostly with the study of logical calculi and their meta-properties. While this may be true for some areas of mathematical logic, this inward turn has been counterbalanced, throughout the second half of the 20th century, by new contacts with other areas, and 'object-level' logical analyses of new notions. The philosophized eschatology of some of my good academic friends just does not match up with what really took place.

A second barrier are divisions of philosophy into *fixed subdisciplines* such as epistemology, metaphysics, ethics, philosophy of language — and why not? — logic. There is some sense to such clustering. But, even in a formal discipline like mathematics, a geography in terms of different countries 'algebra', 'geometry', 'analysis' hides as much as it reveals — and creative mathematicians think in terms of *themes* and *methods* rather than rigid subfields. It is themes and their metamorphoses across subdisciplines that provide the coherence of a field. Here is the worst that can happen. Some atlases of philosophical logic even copy unquestioned philosophical geography (epistemic logic, deontic logic, alethic modal logic), leading to a bad copy of a bad map of reality.

The third trap to be avoided is *system imprisonment* [van Benthem, 1999b]. Many logicians see the structure of logic as a family of formal systems: propositional logic, first-order predicate logic, higher-order logic, modal logics, etc. The formalists may not have won with their philosophy, but their world-view has insidiously inserted itself into the subconscious of the field. This systems view represents a dramatic change from thinking in terms of broad logical themes, such as Negation, Implication, or Validity, which may be studied using particular 'logics', but are not exhausted by them. The formal systems way of perceiving the field has led to a host of 'system-generated issues', which are then touted (often even successfully) as philosophically relevant. Examples include Quine's view of ontological commitment as an issue of ranges for first-order variables, the never-ending discussion about compositionality, or the elusive search for the first-order/higher-order boundary. It is only with great effort that some modern logicians are trying to break out of this mold, realizing that many notions that we study transcend particular formal systems, even though we lack the optimal vocabulary to say what a general 'monotonicity inference', or a general 'recursive definition' is.

A fourth and final barrier to seeing things for what they are is the term *'application'*. Logic is not applied to philosophical problems the way an engineer might apply some technique for computing stress in a bridge. Its role seems much more indirect. Logic offers a technical language with relatively precise meanings as an enhancement of philosophical discourse, and an aid to precise *communication*. This is as useful as having mathematical language around in other disciplines: perhaps not just the Book of Nature, but also the Book of Ideas is written in mathematical language. Also, logical tools can sometimes be used to *analyze arguments* from the philosophical tradition, and throw new light on them. But again, there is usually no counterpart to 'solution' in the engineering sense. The problem gets illuminated: it may go away, but it may also acquire new nuances, and it may even become more urgent than ever! Analysis in logical form also helps bring out *analogies* between different notions or problems that were not observed before. Finally, on the more activist side, logic can sometimes help make philosophical argumentation more precise, and it can help the philosopher *construct new conceptual frameworks*. And yes, sometimes, *meta-theorems* about logical systems in the foundational mode have some bearing on philosophical issues — even though this particular use seems greatly over-rated in practice.

Tracking themes through time

This chapter will not state *a priori* what the role of logic in philosophy is, or should be. Our plan is to just look at some themes as they developed through the 20th century. The richness, and surprising plot twists, of the resulting stories speak for themselves. This way of writing history was tried out in a Stanford course 'Logic in Philosophy' [2003d], and it will be the basis for a new textbook in philosophical logic. Our first two themes show how some of the core ideas of pre-modern logic survived the Fregean revolution, returning in modern forms:

1. Logical form and monotonicity: from syllogistics to generalized quantifiers

2. Reasoning styles: from Bolzano to conditional logic and AI

The next theme shows how formal systems, when viewed with some creative disrespect for authority, are indeed useful 'laboratories' for experimenting with new philosophical ideas:

3. Mechanisms of semantic interpretation: from Tarski to 'dynamics'

Our final two themes address the modern topic of placing actions at logical center stage, relating them to much earlier traditions:

4. Theories and belief revision: from neo-positivism to computer science
5. Dynamic logic, communication, action, and games

In each case, we briefly trace a historical development of ideas up to the present time, without going into details. The references provide further depth, while §9 lists a few further themes. All our story lines involve other fields, mainly computer science. Indeed, confining these things to just philosophy and logic would distort the intellectual history. It would also do a disservice to philosophy, whose themes often cross over into other fields, and then return in new forms — a win-win scenario for everyone involved.

Finally, the references in this paper are meant to be illustrative, but they are not an official historical record. Also, documents mentioned need not be the original sources for a theme described here: they are sometimes later surveys. Indeed, the *Handbook of Philosophical Logic* [Gabbay and Guenthner, 1983] and later extended versions) as well as the *Companion to Philosophical Logic* [Jacquette, 2002] are good general sources.

2 A JOY-RIDE THROUGH HISTORY

For a start, we provide some further historical perspective. To dispel dogmatic slumbers, it is good to realize that there is no stable *ex cathedra* definition of the discipline of logic! Throughout the centuries, there have been many changes in the agenda of logic, its interdisciplinary environment, and the guiding interests of its leading practitioners. The entrenched fundamentalists of today are often the invaders of earlier periods. Classical pre-19th century logic goes back to the dialectical tradition in Antiquity, which realized that success in debate involves fixed patterns of valid, or invalid, reasoning that can be studied as such. For instance, *Modus Tollens* is a wonderful 'attacking move' against one's opponent, by pointing out some false consequence of what she says. Here is a sobering point about human imagination. It took until the mid-20th century before the logician Paul Lorenzen [Lorenz and Lorenzen, 1978] gave the first game-theoretic account of argumentation, doing justice to the fact that the dialectical tradition is about dialogue between different agents instead of lonesome arm-chair deduction, and about winning debates in *interaction* with other participants. Debate was still very much a feature of logical disputations in the Middle Ages [Dutilh-Novaes, 2003], but medieval logic as a whole took logic in the direction of what we would now call ontology and philosophy of language. Logic was then reinterpreted in a more epistemological manner in Kant's famous Table of Categories for the basic forms of judgment. The same august author also proclaimed the 'end of history' for logic — always a dangerous type of prediction...

In the 19th century, leading logicians like Bolzano and Mill saw the subject on a continuum with general methodology, and what we now call philosophy of science. Gradually, influences from mathematics — already present with Leibniz — became prominent, witness the work of Boole and de Morgan on algebraic mathematical methods improving on earlier systems in the syllogistic paradigm. Toward the end of the 19th century, it was no longer a case of mathematics helping logic, but rather the other way around. Frege developed his logical systems for the purpose of securing the foundations of mathematics. Even so, Frege's contemporary Peirce still had a much broader view of what logic is about, including not just deduction, but also inductive and abductive reasoning. It is interesting to see that the 'fundamentalists' won the day with their narrower agenda, making mathematical proof the shining example of logical reasoning — instead of an extreme case of normal reasoning, found only under very rare circumstances. About a century had to elapse before Peirce became a source of inspiration again, now in more computational-logical circles.

Around 1900, modern logic also came to have philosophical impact. Russell's *misleading form thesis* posited that linguistic form, the philosophers' analytical tool until then, can be systematically misleading. One must find the underlying logical forms to express the real meanings and keep argumentation lucid. On this view, logic becomes the 'calculus' of philosophy. Thus, until the 1940s, several logic-inspired programs existed for philosophy. There were the foundations of mathematics with its broad philosophical positions of logicism, formalism, and intuitionism. This research had the Popperian advantage of working with a refutable claim, viz. *Hilbert's Program* for finding conclusive proofs of consistency, completeness, and decidability were around the corner. Gödel's Theorems refuted this, but in the interesting way that mathematical impossibility results normally do. They leave us sadder, but above all: wiser, and full of new directions to pursue. The same was true for Turing's analysis of computation: it showed that many natural computational questions are undecidable, but the fall-out gave us the foundations of computer science. The second, broader line of logic-inspired philosophy is connected with Russell, the early Wittgenstein, and the *Vienna Circle*, a program for a logic-based renewal of methodology, philosophy of science, and philosophy in general. This program, too, had its refutable components, and these were indeed refuted by Quine in 'Two Dogmas of Empiricism' — though Popper later claimed he had helped put in a dagger from behind. Again, the more interesting point are the enduring insights from this second logical line, such the seminal work by Carnap, Reichenbach, and others on methodology and meaning.

These golden years produced the basic *modus operandi* in modern logic, of unquestioned importance even to those who would now like to broaden the agenda again. First-order logic, the modern notion of a truth definition and semantic validity, completeness theorems, proof theory, and many other basic notions go back to work in the foundationalist logic of those early days. Over time, the broader philosophical repercussions of these notions and insights about them started penetrating, witness Tarski's semantic conception of truth, or Turing's analysis of computability and intelligence.

Nevertheless, some originally 'logical' topics did drop by the wayside, as communities started forming with their own subcultures. Carnap is now considered a 'philosopher of science' — and during the 1950s philosophers of science took care of topics that had no place on the foundationalist agenda, such as inductive reasoning, explanation, or scientific theory structure. At the same time, the 1950s also saw the first challenges to the logical way of doing philosophy. Of course, mainstream philosophers perhaps ignored logic, and only a few traditionalist logicians were still bewailing the Fregean Coup — sometimes for very legitimate reasons [Wundt 1880/3; Blanshard, 1964]. But now, criticism came from within the analytical tradition, with the Wittgenstein-Austin 'natural language turn'. Informal linguistic analysis could serve just as well for the purposes of philosophy, or, as Austin said about common usage: "There is gold in them there hills". New appealing paradigms such as 'language games' arose, and there were even attempts to replace logic in philosophical curricula by courses in *game theory*, the upcoming formal paradigm of the day. Just try to imagine another possible world, where Wittgenstein, Nash, and Aumann are the heroes of formal philosophy!

But there was enough vigor in the logical tradition to survive all this. Indeed the 1960s/1970s became the heighday of an emergent milieu called *philosophical logic*, providing new formal tools for looking at long-standing philosophical problems. This was often triggered by taking a much closer look at natural language and our natural reasoning practices than what had been usual in the foundationalist tradition. It suffices to mention a brief list of eminent authors emerging in that period, whose work has sparked a lively follow-up which continues until today: Hintikka, Prior, Geach, Kripke, Rescher, Lewis, Montague, and Stalnaker. Their work was concerned with what may be called *modal logic* in a broad sense: including time, knowledge, duty, action, and counterfactuals. Many of these logicians wrote influential publications which started philosophical research lines, witness famous book titles like *Knowledge and Belief*, *Naming and Necessity*, *Counterfactuals*, or *Formal Philosophy*.

Incidentally, around 1970, logic was heavily attacked inside *psychology* in [Wason and Johnson-Laird, 1972], a publication which convinced many psychologists that logic was spectacularly unsuited for describing the human reasoning with which we really solve our problems and get by in life. Not many logicians will have cared about this threat in the distance, especially since they may have felt that psychologists were not quite people one would want to associate with anyway, given Frege's famous 'anti-psychologism'. But there may be more to this in a modern light: see below.

At the risk of offending some, I would say that the Golden Age of philosophical logic was over by 1980, and this new stream entered into steady state, just as mathematical logic had done, with technical depth taking over from the original intellectual passion. But at the same time, the flame of logic carried over to other fields. Logic gained importance in disciplines beyond philosophy and mathematics, such as linguistics, computer science, AI, and economics — and these are still flourishing interfaces today, often with their own communities, publications, and gatherings. Many topics from the mathematical and philosophical heartland acquired new lives in these settings, witness new theories of meaning, information update, belief revision, communication, and general interaction. One of the most surprising developments has been the design of logic systems of programming, performing various intelligent tasks (cf. the program of 'Logical AI' [McCarthy, 2005], or Reiter's [2001] book *Knowledge in Action.* Even if logics do not completely capture all of our natural cognitive habits, they may at least be consistent with them, and then — gently inserted — enrich our repertoire of behavior! Thus, contemporary logic finds itself in a much more diverse interdisciplinary environment, perhaps closer to the richer agenda of logicians from earlier centuries.

At the same time, challenges persist. In both computer science and game theory, and indeed philosophy itself, *statistical* and probabilistic methods have gained importance, dealing with phenomena where logic has little to offer. Indeed, modern evolutionary approaches to games and methodology rely on dynamical systems rather than logic. Combining logic and probability seems a major challenge today. An even aggressively anti-logical stance surfaced in the computational paradigm of massive parallellism, replacing the Turing machine view of cognitive behavior. At the central AI conference on Knowledge Representation in Toronto around 1990, a keynote speaker felt we were like the French Ancien Regime in our ballroom, while teeming masses of neural netters were shouting "A la laterne, les aristocrates". But again, history has its own ways. By now, it is becoming clear that neural nets and logics can live in harmony [d'Avila Garcez *et al.*, 2002; Leitgeb, 2004; Smolensky and Legendre, 2005], and that dynamic logics and dynamic systems are a match. Thus, the current interface between logic and cognitive science is in flux again [Hodges, 2005], and doom scenarios have not materialized — so far.

Role models

Another way of high-lighting an intellectual development looks at typical representatives. A Dutch professor of linguistics once produced, when challenged about a logical turn in his work, the following syllogism: "Linguistics is that work which is being done by prominent linguists. I am one of this country's most prominent linguists. Therefore, what I am doing is linguistics". Role models for a person-oriented line would include philosophers/logicians like Peirce, Ramsey, Carnap, Geach, Dummett, Hintikka, Grice, Kripke, Lewis, Stalnaker, Belnap, Kamp, and Barwise — and a very interesting chronicle might result. What did these leaders consider the most pressing themes, and why? Nevertheless, in the rest of this paper, we will consider themes rather than people.

3 LOGICAL FORM AND NATURAL LOGIC: FROM SYLLOGISTICS TO GENERALIZED QUANTIFIERS

Natural logic

Is natural language unsuitable as a vehicle for logical inference? Frege thought so, and cited its 'prolixity' as a reason for designing the formal language of *Begriffsschrift.* In opposition to this, a tiny band of traditionalist critics defended classical syllogistics as staying much closer to natural inferences that people actually make [Sommers, 1982; Englebretsen, 1981]. This debate was reopened in the 1980s under a foreign influence, viz. the penetration of logic into *linguistic semantics.* Richard Montague's earlier work [Montague, 1974] — influenced to some extent by Donald Davidson's critical stance toward the Tarskian formal logic of his day — was seminal in this development, in being two-edged. It used formal logic to analyze natural language expressions — but in doing so, it also showed that natural language did have a respectable structure that supports impeccable meanings and inferences. The modern theory of *generalized quantifiers* spawned by Montague Grammar has replaced Misleading Form antagonism by concrete questions about the actual interactions of reasoning encoded in natural language and formal logic.

Traditional logic

A brief account of logical form runs as follows (cf. GAMUT [1991]). Traditional logic used the subject predicate format S is P, reflecting the structure of natural language sentences. Even so, this was not really a logic of sentence forms, but of intensional relations between concepts denoted by the terms 'S, P' — as is still clear in Kant's famous Table of Categories. The inference engine for this reasoning about concepts was Aristotle's syllogisms for the basic quantifier expressions *all, some, no, not all,* forming the classic Square of Opposition. In the Middle Ages, the coverage of syllogisms was extended by the ingenious doctrine of *distribution.* This uses the fact that quantified premises allow for inference in deep sentence position, providing we use universal relationships to replace only occurrences of predicates that are 'about all objects' (in modern terms: downward monotonic ones), or their positive monotonic counterparts. This allows inferences that go beyond the traditional syllogism, such as

> If all men are mortal, and no rich mortal's son escapes Judgment Day,
> then no rich man's son escapes Judgment Day.
> If all men are mortal, and Judgment Day is hard on many men,
> then Judgment Day is hard on many mortals.

In this way, traditional logic could and did deal with many forms of judgment and reasoning — and even Leibniz felt it needed perfecting, rather than demolishing.

The critics win the day

In the 19th century, progressive waves of critics stormed the bastion. Boole's attack was implicit, in showing how at least propositional reasoning could be dealt with in a perspicuous algebraic manner owing nothing to Aristotle. Later historians found analogues in the work of the Stoic logicians — but it was Boolean insights that rehabilitated Stoic logic [Mates, 1953], rather than Stoicism supporting Boole. Famous examples of the 'inadequacy' of traditional logic still taught today include De Morgan's non-syllogistic forms:

> All horses are animals. *Therefore*, horse tails are animal tails

This purported knock-down objection is actually well within the scope of the medieval theory of Distribution. The fact that a binary relation occurs ("animal x possesses tail y") is completely immaterial, as the modern equation of the Syllogistic with a fragment of *monadic* predicate logic is a highly uncharitable later construal. Frege's later examples of the inadequacy of traditional logic concern more essential uses of relations:

> from "The Greeks defeated the Persians at Plataeae,
> to "The Persians were defeated by the Greeks at Plataeae".

Like most conservatives, Frege preferred high-lighting the aristocratic Spartan victory at Plataeae over the prior democratic Athenian one at Salamis. But his logical point is a strong one. Moreover, his anti-intensional compositional treatment of quantifiers (see the compelling story of Dummett [1973]) was a startling innovation. A sentence like "Everybody walks" now stood for a multitude of individual instances, such as "Mary walks", "John walks",... This extensional view derived complex sentence meanings recursively from properties and relations over some given domain of individual objects. And like all creative simplifications in science, it paved the way for new insights and concerns.

Why the quick collapse?

It is an intriguing question why the traditional bastion crumbled so fast. Some modern historians (cf. [Barth, 1974]) tried to paint an image of an Old Guard defending itself to the bitter end, and evil forces pitted against the Russelian and Wittgensteinian forces of Light (cf. the 'Kingdom of Darkness' in [Geach, 1972]). But what they mainly managed to drag into the open were some pitiful traditional logicians, looking bewildered as the walls were tumbling around them. Indeed, the complaints of the old guard about unfairness were often on the mark. Van Benthem [1981] discusses two so-called 'crude errors' of traditional logic: its reduction of binary predicates to unary ones, and its use of generic objects, and shows how both still make a lot of sense, when viewed with some mathematical sensitivity. Still, the bastion did crumble, largely because the critics had a more dynamic agenda of research — and also, because bastions crumble, not so much because of the vigor of external onslaughts, but because of the internal dry-rot of stagnation and boredom.

From misleading form to Montague's Thesis

As we saw, the Misleading Form Thesis made logic a favored calculus for philosophy. Even so, natural language refused to go away in philosophical practice and theory. Indeed, the very success of Montague's work around 1970 on logical grammar and semantics of natural language cast the issues in a completely new light. Montague retained the subject-verb form of natural language sentences, with its basic pattern $S \Rightarrow NP\ VP$. To do so, he treated noun phrases as *generalized quantifiers*, i.e., second-order properties of predicates. Thus, "Every bird sings" says that the property of singing holds for all birds — and to make "Tweety sings" come out in the same way, an individual bird like Tweety was 'lifted' to become the set of all her properties. This generalized quantifier view can also deal with iterated quantified expressions, as in "Every bird sings a song". The success of this approach motivated what became known as Montague's Thesis:

> 'There is no difference of principle between natural
> and formal languages'.

This is the Misleading Form Thesis reversed! From then on, the story shifted to linguistics, where Montague semantics became a powerful paradigm, and even computer science [Janssen, 1981; de Roever *et al.*, 2001], as programming languages, too, allow for the same sort of compositional logical treatment.

Fine-structure: generalized quantifier theory

The *NP VP* format for quantified *NP* expressions has a three-part structure *(Det A) B*, where *Det* stands for a *determiner* expression: "(*Every* bird) sings". Thus, quantifiers themselves may be viewed a binary relations between extensional predicates A, B. E.g., *every* stands for inclusion between the sets of "birds" and "singers", and *some* for overlap. This account also works for non-first-order quantifiers. E.g., *most AB* says that the size of the set $A \cap B$ is greater than that of the set $A - B$. With this view, many new issues have come to light over the past two decades [Keenan and Westerståhl, 1997; Peters and Westerståhl, to appear]. Of relevance to us here is the inference pattern of *monotonicity*, describing stability under shifts in predicates. E.g., a true statement *every AB* remains true after replacement of its A-argument by a predicate denoting a subset, and after replacement of its B-argument by a superset. In technical terms, *every* is downward monotone in A, and it is upward monotone in B. The other classical quantifiers in the Square of Opposition exemplify the other three types: e.g., *some AB* is upward monotone in both its arguments. Modulo some mathematical assumptions, this can even be reversed: making the classical quantifiers the only 'doubly monotone' ones (a theorem in [van Benthem, 1986]).

Monotonicity reasoning

Now, it turns out that this monotonicity behavior can be described completely generally for any occurrence of a linguistic predicate inside any linguistic expression, simply by interleaving it with grammatical sentence construction in some logic-friendly paradigm like *categorial grammar* [van Benthem, 1986; Sanchez Valencia, 1991; van Benthem, 1991]. This categorial 'monotonicity calculus' has been rediscovered several times over the past few years for its elegance and naturalness, and with hindsight, it can be discerned in the traditionalist manifesto [Sommers, 1982]. Today, it even serves as a platform for cognitive experiments [Geurts and van der Slik, 2004]. This vindicates the medieval doctrine of Distribution, and it accounts for perhaps the majority of the 'fast inferences' that language users make. Indeed, similar systems of light but pervasive quantifier inference also turn up in computer science, with fast algorithmic tasks in handling data bases.

Co-existence of natural logic and formal logic

Monotonicity is just one form of 'natural reasoning'. Other principles include systematic restriction of ranges for quantifiers (already described by Peirce), or 'context shifting' across time and space (cf. [van Benthem, 1987]). A complete system of all aspects of natural logic still remains to be formulated. Even so, it seems clear that first-order logic is by no means the only inferential standard. It rather seems a sort of formal laboratory for defining and studying certain notions (including monotonicity), which we can then use to describe actual inferential phenomena, if only in the form of a contrast. The general idea behind this view also reflects current computational realities — as computer scientists populate the world with an ever-growing arsenal of semi-formal languages, artificially designed but used in practice. There is a fruitful *co-existence* of natural and formal languages, and their mutual influences are probably the best way of laying earlier all-or-nothing discussions about misleading form, or about the 'impossibility' of a systematic theory of meaning for natural language, to a well-deserved rest.

But what about philosophy?

The moral of this story is that the traditional debate about wholesale choices between natural or formal language as the philosopher's tool are misguided. Logical languages are an *enhancement* of philosophical practice — and in that light, philosophy is precisely one of the mixed settings that we advocated.

4 REASONING STYLES: FROM BOLZANO TO CONDITIONAL LOGIC AND AI

Styles of reasoning

Most textbooks define 'validity' of an inference once and for all, as truth of the conclusion in all models where the premises hold. I still remember my flash experience as a student finding Bolzano's *Wissenschaftslehre* [1837] in our library, and seeing how, long before Tarski, someone had looked at what logic is about, and made some crucial decisions differently. Bolzano's best-known notion of logical consequence is indeed like Tarski's [van Benthem, 1985], be it that *nothing*, rather than everything, follows from inconsistent premises. But beyond that, Bolzano describes the aim of logic as charting the *varieties of reasoning styles* that we apply to different tasks. These are deductive in mathematics, often inductive in daily life, or — in his view – strictest of all in philosophical argumentation, where drawing *relevant* conclusions is of the essence. Bolzano then lists formal properties distinguishing such reasoning styles. E.g., he points out — in modern terms — that some are *monotonic* (adding new premises does not affect earlier conclusions), while others are not. This idea of a variety of reasoning styles returns with Peirce, the neo-positivists, and again in modern *AI*, when nonmonotonic logic arrived on the scene in the 1980s as an account of how people actually solve puzzles and go about practical planning tasks [McCarthy, 1980]. It was only in the 1990s that what may be called *Bolzano's Program* came into its own with the broad spectrum of nonmonotonic reasoning styles that we know today, based on a wide variety of motivations, including default reasoning, resource-sensitive reasoning, and neural computation. And we still have not explored its full depth, which also includes a more dynamic view of the role of changing vocabulary in reasoning [Rott, 2001b; van Benthem, 2003b].

Logic and methodology

Bolzano's book combines what we would call logic today with methodology of science. This was quite common in earlier days. Tarski's own basic textbook of the 1940s was called *Logic and Methodology of the Exact Sciences*, and when Tarski and Beth started world conferences in what they saw as their field [Van Ulsen, 2000], they initiated the still existing tradition of joint meetings in "Logic, Methodology, and Philosophy of Science". In the meantime, the different component communities have drifted apart — but there are also confluences again, partly through the intermediary of Artificial Intelligence.

Different functions of reasoning

Valid semantic consequence $\phi \models \psi$ in predicate logic says that all models of ϕ are models for ψ. Or, in terms of proof, $\phi \vdash \psi$ says that there is a derivation consisting of elementary steps from ϕ to ψ. By Gödel's completeness theorem, the two notions are extensionally equivalent for first-order logic, and the same holds for many other logical systems. Either way, consequence can serve very different reasoning purposes. In a forward direction, one derives new conclusions, like a mathematician proving new theorems. But in a backward direction, one refutes a hypothesis by deriving some false conclusion: a process of hypothetical reasoning and falsification. A first connection with methodology here is Popper's insistence on the backward process of refutation as more fundamental than accumulating truth. Variety of functions even increases when we consider actual reasoning practices. E.g., in a juridical procedure, the prosecutor must show that a given assertion, guilt of the accused, follows from the evidence on the table. But in the same setting, the lawyer must show that the accused's innocence is consistent with that evidence. Thus, inference is deeply intertwined with 'consistency management'. Perhaps the richest view of logical tasks is found with Peirce as early as the 19th century. His distinctions between *deduction* (forward-looking derivation), *induction*, and *abduction* (backward-looking search for premises for given conclusions) fit remarkably well with the many inferential processes that humans engage in [Aliseda-Llera, 1997; Flach and Kakas, 2000].

Hypothetico-deductive explanation

With this broad view of uses, the methodology of science becomes relevant at once, with its further natural varieties of reasoning such as *explanation* or *confirmation.* These notions occur in science, but they also resonate with ordinary practice. A concrete case is the Hempel–Oppenheim view of explanation. Given a theory T, certain facts F explain a given observation O if

$$T + F \models 0,\ \ \mathit{not}\, T \models 0,\ \ \mathit{not}\, F \models 0.$$

This is the sense in which the lighting of a match in a gasoline-saturated garage explains the explosion of the car, given the laws of physics. This simple formulation still suppress a third ingredient in the inference, the 'auxiliary hypotheses' which make sure that the situation described is a 'normal one', fit for standard explanations. E.g., we assume that oxygen is present in a normal garage, and that the car does not run on wood. The new notion is no longer classical consequence. In particular, it is *nonmonotonic*: stronger premises F may no longer explain O in T, as they imply O by themselves — and stronger theories T may no longer need F for deriving O. What is interesting here from a logical point of view are two things. We work with structured premises, whose parts play different roles. And we mix logical consequence with non-consequence. But is there any logic here? And is the notion of interest? Carnap worried that any fact O with an explanation F in T also has the trivial explanation $T \to F$. So, is genuine explanation an art beyond logic? At least there is more to the pattern of explanation.

As for other notions studied in methodology, we note that Hempel [1965] explained confirmation of a law as truth of the lawlike proposition in a *minimal model* of the evidence so far. This is the sense in which successive observations

$Qd_1, Qd_2, \ldots$ would confirm the universal regularity that $\forall x Qx$, even without logically implying it. Again we see a divergence from classical consequence, as we now look at just the smallest models of the premises, rather than all of them, to check for the conclusion.

Counterfactuals

One very fruitful strand in the philosophy of science has been the analysis of counterfactual conditionals. In thinking about what makes a statement a *law*, as opposed to just an arbitrary generalization, Goodman observed in the 1940s that a genuine law supports *counterfactual assertions* about situations that did not actually occur, such as, "if this match had not been lit, no explosion would have occurred". Carnap observed in the same vein in the 1950s that scientific theories with laws often support *dispositional* predicates, such as 'fragile': a fragile object 'would break if hit', even if it stays whole forever. Counterfactuals are clearly unlike ordinary logical conditionals. In particular, again, they are nonmonotonic. If this match had not been lit, but a hand-grenade was thrown, an explosion would have occurred after all. By the 1960s, the time was ripe for a generalization behind these notions.

Conditional logic

An early appealing account of nonclassical conditionals $A \Rightarrow B$ comes from Ramsey [1931]. It says this: "add A to your stock of beliefs T, make the minimal adjustment to T to keep the addition of A consistent, and then see if B follows." At the time, this seemed far outside of the logical tradition, but things changed in the 1960s when Lewis and Stalnaker initiated conditional logic. Here is their basic idea [Lewis, 1973]. Assume that situations or models come ordered by some relation $\leq$ of *relative similarity.* Its source can be a variety of considerations: relative plausibility in one's judgment, objective likeness, and so on. Now we say that

$$A \Rightarrow B \text{ is true at } s \text{ iff } B \text{ is true in all } A\text{-worlds } \leq\text{-closest as seen from } s.$$

This scheme generates a true logic of valid principles for reasoning with this notion. These include Reflexivity, Conjunction of Consequents, Disjunction of Antecedents, and Upward Monotonicity for Consequents. What fails is, of course, Downward Monotonicity for Antecedents. Instead, on a minimal view of the comparison ordering $\leq$, the only thing that does hold for conditional reasoning is a substitute, called Cautious Monotonicity:

$$A \Rightarrow B, A \Rightarrow C \qquad \textit{implies } A\&B \Rightarrow C.$$

This phenomenon of 'modified rules' would become significant in the 1980s. Conditional logics have been used for many purposes in philosophy, including Lewis' own counterfactual account of causality, and the counterfactual account of knowledge in Nozick [1981]. In each case, counterfactual statements express the robustness in inference that a natural epistemic notion like ‘law’, ‘cause’ or ‘knowledge’ is supposed to support.

Nonmonotonic logic in AI

The idea that reasoning styles depend on the task at hand returned in strength in *AI* around 1980. McCarthy [1980] argued that our ordinary processes of problem solving make all sorts of systematic additional assumptions, in particular, the use of *minimal models* (in some suitable sense) for the given premises, representing the fact that we are assuming a ‘closed world’ setting for the task at hand, without surprises. Thus, in *circumscription* (cf. [Shoham, 1988]), we say that

> the conclusion must hold in all *minimal models* of the premises.

Note the similarity with Hempel’s view — though Circumscription is much more powerful and sophisticated. Other examples of such styles have come up ever since, including abduction in expert systems and logic programming. Thus, at last, Bolzano’s pluralistic view was vindicated, be it not by logicians and philosophers (yet), but by computer scientists. Here ‘minimality’ is one of the constants of research into logics for a variety of reasoning tasks. It returned in accounts of practical *default reasoning* (defeasible rules of thumb; [Reiter, 2001; Veltman, 1996; Shanahan, 1997], and again in the theory of *belief revision* [Gärdenfors, 1987], which will be the topic of §6.

Structural rules

Summarizing two decades of research on all the new-fangled logical systems that have been proposed for dealing with circumscription, default reasoning, and belief revision, it seems fair to say that the core logic of minimality *is* conditional logic. Bolzano’s observations about formal properties of different reasoning styles then have their correlate in the deviant properties of classical consequence versus that expressed in conditional logic. The locus of these differences is often put in the so-called *structural rules* governing the over-all properties inference without any special logical operators. Complete packages of structural rules have been found for many new notions of inference, including Bolzano’s own [van Benthem, 1996a; 1996b; 2003b; 2003e]. Deviant structural rules also occur for reasons very different from minimality, however, witness their role in relevant logic, and in the thriving area of ‘substructural logics’ of syntactic combination, dynamic update, or interaction [Dosen and Schroeder-Heister, 1994].

Tricky issues

Given all this, is there just proliferation, or can we now begin to chart a neat landscape of natural reasoning styles? No simple map has emerged yet, and some people feel the only way of constraining the options is by asking some external authority like cognitive science which nonmonotonic logics match realities in the mind, or the brain. Another problem is this. Notions that deviate from classical consequence are often more complex in their cost of computing validity or satisfiability. On top of predicate logic, they often become non-axiomatizable, or worse. This seems strange, given their motivation in ordinary practice — and the phenomenon is still not totally understood.

Migrations

Even so, our story across logic, philosophy of science, and computer science shows that topics like reasoning styles and conditionality migrate across borders without any problems. This continues until today. Van Rooy and Schultz [2004] apply circumscription to the semantics of questions in linguistics (cf. [van Benthem, 1989]), suggesting that the natural logic of §3 may also include nonmonotonic mechanisms in information exchange. More generally, through these migrations the distinction between scientific method and common sense — once thought so iron-clad — is evaporating. There is just one form of rationality, whether displayed in our home kitchen or the Halls of Minerva. And often, it is computer science that brings the viewpoints together in one single perspective [McCarthy, 2005]. Indeed, computer science plays a role in the development of about every logical theme in the last century. And perhaps this is not so strange. When all is said and done, computer science and its more pregnant form of Artificial Intelligence are just — in Clausewitz' happy phrase — 'the continuation of philosophy by other means'. And the relative neglect of computation and information as fundamental categories in mainstream philosophy (cf. [Floridi, 2004]) mainly seems a case of rejecting that which is closest to us.

5 MECHANISMS OF SEMANTIC INTERPRETATION: FROM TARSKI TO 'DYNAMICS'

First-order predicate logic may not be a very realistic description of our natural logic, but it is an amazing model for studying a tightly interwoven set of issues at the same time. The formal language has a clear inductive grammatical structure, and the mechanism of semantic interpretation capitalizes on this in a compositional manner. Thus, it explains the most complex forms of assertion by repeating a small number of elementary truth conditions for separate operators. This setting made it possible to study issues of definition and expressive power in logical model theory. This format of interpretation has been very influential in philosophy, influencing theories of truth through Tarski's classic "Der Wahrheitsbegriff in den Formalisierten Sprachen" from the 1930s (cf. [Tarski, 1956]), and innovative later extensions by [Davidson, 1967; Kripke, 1975], and others. Even so, the laboratory is still capable of generating new ideas!

Standard semantics

The key notion in first-order semantics is that of truth of a formula ϕ in a model $\mathbf{M}$ under some variable assignment s:

$$\mathbf{M}, s \vDash \phi.$$

The key compositional clause uses the assignment in an essential manner:

$$\mathbf{M}, s \vDash \exists x\phi \qquad \text{iff} \qquad \text{for some object } d\mathbf{M}, s[x := d] \vDash \phi.$$

In this way, pulling apart the different ingredients in this scheme, truth is a relation between an expression in a formal language and some language-independent structure of objects with their properties and relations. The relation is mediated by means of two links. One is an *interpretation function I* assigning fixed denotations to predicate letters and other expressions whose meaning is considered constant in the current setting. The other is the *variable assignment s*, providing an auxiliary denotation for variable parts of the expression, which may change in the process of evaluation. Logicians look at this scheme from various angles. Given a formula ϕ, one can study its models $MOD(\phi)$. Conversely, given a model $\mathbf{M}$, one can study its theory $Th(\mathbf{M})$: the sentences made true by it. Or, less common but quite possible, given a formula and a structure, one can study interpretation functions making the formula true in that structure.

Basic features of the set-up

This scheme embodies several historical steps, and nontrivial conceptual issues. Its decoupling of formal syntax and semantic evaluation is a major abstraction, as our natural language wears its interpretation on its sleeves, leaving only some freedom in interpreting pronouns and other variable 'lice of thought' — as Italo Calvino once described them. Next, the notion of a model brings its own issues, as first-order models are a *representation* of reality, not reality itself — except perhaps in some parts of abstract mathematics. Indeed, decades before Tarski, de Saussure proposed his famous Semiotic Triangle of *language, representation, reality.* First-order semantics ignores the latter — but issues of 'fit' to reality do not go away. Indeed, a three-part scheme of interpretation may be found in the work of Church and other logicians. With much more impact, this perspective returned in the 1980s, with the work by Kamp and Heim on discourse representation theory (cf. the survey in [van Eijck and Kamp, 1997]). Of the many further features of the scheme, let us just note the compositionality once more. Dummett [1973] called this Frege's major insight — though modern scholarship has cast some doubt on Frege's paternity for this principle. Instead, he did place a lot of emphasis on a principle of *Contextuality*, stating that meanings of expressions are never to be studied in isolation from their context. Hodges [2001] is a delightful analysis of the connections between the two principles.

Philosophical and linguistic aspects

An amazing amount of philosophical literature is tied directly to the logical semantics for which the above scheme is a paradigm. Some classics are [Davidson and Harman, 1972; Putnam, 1975; Field, 1972], while more modern perspectives are in [Etchemendy, 1990; Zalta, 1993]. Some of these philosophical issues returned more concretely in the semantics of natural language, such as Montague's pioneering work that was mentioned in §3. E.g., concerns with compositionality haunt that literature until today, with discussions fed by the study of concrete 'non-Fregean' quantifier combinations that resist easy iterative decomposition [Keenan and Westerståhl, 1997]. Also, the status of the models becomes more pregnant here, as we need to take ontological decisions. An unresolved case is the ubiquitous co-existence of 'count' and 'mass quantifiers', and thus, the interplay of discrete and continuous objects in discourse and reasoning.

Proliferation of frameworks

In the 1980s and 1990s, the reign of classical semantics was challenged by a host of new semantic frameworks. *Situation semantics* [Barwise and Perry, 1983] emphasized that meaning involves many situations, rather than just one model $\mathbf{M}$: including a situation of utterance, a described situation, and a resource situation. Thus, a true account of semantics involves a network of small contexts. This perspective is coming to the fore these days in many disciplines. Modern interdisciplinary conferences on context [CONTEXT, 2005] bring together philosophers, linguists, and computer scientists with overlapping agendas. Next, the above-mentioned *discourse representation theory* is still a major paradigm for meaning in linguistics and computer science, as discourse representation structures are the natural format for computational processing. In particular, the emphasis now shifts to the dynamics of interpretation: we understand a sentence, not by evaluating it in some given model, but rather by constructing some sort of representation of the described situation. Dynamics of interpretation is also crucial to other computational paradigms, such as 'parsing as deduction' [Pereira and Warren, 1983], or 'interpretation as abduction' [Hobbs *et al.*, 1990]. The *Handbook of Logic and Language* [van Benthem and ter Meulen, 1997] documents this amazing outburst of semantic creativity, little of which has yet penetrated to philosophical discussions of truth and meaning. There are even other new approaches — of which we discuss one in a little more detail. It demonstrates the proper use of the 'first-order laboratory': not as a source of orthodoxy, but as a source of dissent.

Dynamic semantics

An idea common to many newer semantic frameworks is that semantics is really about *dynamic procedures* of evaluation, rather than just static relationships between formulas and models. Can we make these actions themselves an explicit part of the logical apparatus? A well-known proposal of this sort is *dynamic predicate* *logic* (*DPL*; [Groenendijk and Stokhof, 1991]). The motivation is provided by the phenomenon of anaphora, i.e., how we give coherence to what we say. We study this in a simple first-order laboratory setting. In the grand tradition of 'misleading form', students are usually taught some translation folklore to make first-order formulas fit natural language forms. Consider the following statements:

1 *A man came in. He whistled.* The two underlined phrases can co-refer.
2 *He whistled. A man came in.* The two underlined phrases cannot co-refer.
3 *If a man came in, he whistled.* The two underlined phrases can co-refer.

A direct translation $\exists x\ Cx \& Wx$ for 1 does not give the intended scope, and so we use a bracketing trick to get $\exists x(Cx \& Wx)$. A translation $Wx \wedge \neg\exists x Cx$ for 2 does give the right scope, and hence no tricks need be used here. But again, a direct translation $\exists x Cx \rightarrow Wx$ for 3 does not work, and so students are taught a bracketing trick plus a quantifier-change to $\forall x(Cx \rightarrow Wx)$. This seems terribly unprincipled and ad-hoc. By contrast *DPL* changes the mechanism of first-order semantics, reading formulas ϕ as *procedures*, whose meanings are transition relations between variable assignments s, viewed as temporary states of the evaluation process. This can be done entirely compositionally, without changing either the first-order language or its models. This illustrates the more complex semantic world since Montague, where ideas can travel between philosophy, linguistics and computer science. The procedural view comes from the semantics of imperative programming languages. In that setting, expressions in programs are instructions for successive changing of states, viewed as assignments of data-objects to variables. Think of registers filled with transient objects. Executing an imperative instruction $x := 2$ replaces the current content of register x with a value 2. More precisely, *DPL* records traces of successful executions in the following format:

> $\mathbf{M}, s_1, s_2 \models \phi$ iff there exists some successful evaluation of ϕ in $\mathbf{M}$ starting from input state s_1 and ending in output state s_2

The semantic clauses treat atomic formulas as tests on the current assignment, conjunction becomes composition of actions, negation is a failure test, and the real dynamics of state change is in the clause for the existential quantifier, which looks for a witness:

$$\mathbf{M}, s_1, s_2 \models \exists x \text{ iff } s_2 = s_1[x := d] \textit{ for any object } d \textit{ in the domain}$$

When read in this dynamic way, the direct translations of all three examples above come out just as they should without re-bracketing tricks. While this may look like just a new way of doing natural language semantics, the scheme really challenges many presuppositions in logic and philosophy. For instance, meaning and inference now reflect an *algebra of procedures.* E.g., the scope widening in $\exists x Cx\ \&\ Wx$ is just *associativity* of composition. Or with a *universal quantifier* $\forall x\phi = \neg\exists x\neg\phi$, evaluation procedures for $\exists x Cx \rightarrow Wx$ and $\forall x(Cx \rightarrow Wx)$ turn out algebraically equivalent. Indeed, *DPL* comes with its own notion of dynamic consequence in which processing the successive premises validates the conclusion. This is again a nonstandard, and nonmonotonic reasoning style (cf. §3) — for which we refer to the literature.

More radical dynamics: interactive games

The general idea of dynamic semantics is that meaning involves *information change*, an idea also known in discourse representation theory as 'context change potential'. A more radical version of this is *game-theoretic semantics* [Hintikka and Sandu, 1997]. Now, evaluation is no longer just a procedure for one 'checker', but a *two-person game* between a Verifier and a Falsifier with opposing roles concerning the given formula. The former tries to show that it is true, the latter that it is false. In that way, *interaction* becomes essential to understanding meaning and inference. In his many publications on the subject, starting from the 1960s, Hintikka has spelled out a wide range of philosophical implications of taking this stance. Even more radically game-theoretic views of interpretation, partly along the lines of Grice [1975], are coming up these days, as semantic research focuses on *interpretative equilibria* sought by speakers and hearers in actual communication [Lewis, 1969; Parikh, 2001; van Rooy 2004].

Languages again

Contemporary semantic frameworks involve ideas from logic, linguistics, and computer science — again demonstrating the unity of perspective across natural, formal, and programming languages. What we have said so far by no means exhausts this potential. For instance, [van Benthem, 1996; 1999a] proposes yet another procedural reinterpretation of first-order semantics! It uncovers a decidable modal base logic of compositional evaluation, which includes all the monotonicity reasoning of §3. On top of this, further validities express geometrical constraints on the flow patterns of assignment change. The undecidability of first-order logic, often taken as an inevitable hallmark of reasoning with quantifiers, then becomes a side-effect of working with special computation spaces. In the light of these modern developments, many philosophical discussions of first-order logic and the import of its classical meta-theorems seem outdated, and hopelessly conservative.

An interesting issue is what all this says about our understanding of natural language, the vehicle of both common sense and technical philosophy. Changing information states of readers and hearers is definitely a major purpose of any type of discourse, including philosophical dialogue. If we take this dynamic perspective seriously, some people feel it may be the computer science perspective that should have the last word: by viewing natural language as *the programming language of cognition.* Or, if this slogan is too crude for serious philosophy, here is another way of phrasing the point of this section. Meaning is not just *how things are*, it also essentially involves — as Wittgenstein already said — *what we do.*

6 THEORIES AND BELIEF REVISION: FROM NEO-POSITIVISM TO COMPUTER SCIENCE

The dynamic turn

A computational stance involves a *tandem view.* Representations should always be constructed with some process in mind, while processes should be designed with some data structure in mind. In a sense, this happened even in the early days of logical analysis (cf. §4). After all, words like 'explanation' denote a dynamic activity as much as a static proposition or relation between propositions. One conspicuous development over the past few decades has been a Dynamic Turn [van Benthem, 1996], moving beyond static relations between propositions on the pattern of logical consequence to activities like proving, testing, learning, or changing one's mind, as independent objects of study. In this perspective, traditional interfaces between logic and philosophy acquire a double aspect: what are the activities involved, and what representations do they work with? The dynamic semantics of §5 is one concrete illustration of such a move, but there are many others. In this section, we apply this style of thinking to the scientific theories that briefly surfaced in §4.

Theory statics

The simplest logical view of a theory is just as a set of sentences. This, of course, will not do as an account of any plausible notion of theory in the empirical sciences — and not even as an account of the 'structured opinions' which people have in the realm of common sense. Richer views of theories developed in the philosophy of science, witness the work of Ramsey [1931], Beth [1949], and Przelecki and Wojcicki (cf. [Przelecki, 1969]). There is a lucid summary of all this in the Introduction to Suppe [1977], which covers both syntactic views in the neo-positivist tradition, and more semantic ones in the style of Suppes and Sneed. Kuipers [2000] is an up-to-date exposition of modern notions of theory, including an account of the major *inter-theory relations* that have been of interest to philosophers, such as various kinds of extension and reduction. Theory structure is an excellent theme by itself as a running interface between logic and philosophy (cf. [Pearce and Rantala, 1984; van Benthem, 1982]). In line with all our topics so far, it has surprising connections to computer science, in particular the area of 'abstract data types' [van Benthem, 1989] — but we forego this particular story here. For, theories by themselves are just static representations of concepts and information at the disposal of individuals or communities. In line with our main interest, we want to see how such theories grow, and more generally: *change.*

Theory dynamics

Some of the most spectacular philosophy of science is about ways in which theories change. Sometimes, there is cumulative piece-meal growth, but occasionally, there are more drastic paradigm shifts, as described in Kuhn [1962]. Processes of theory change were made into serious objects of logical study in the work of [Alchourron *et al.*, 1985; Gärdenfors, 1987; Rott 2001a]. The simplest process is *update with consistent new information,* obtaining a new theory $T + \phi$. This adds the new formula ϕ to the current theory T, and then closes the result deductively. Alternatively, viewing the theory as a set of models *MOD(T),* it forms the intersection *MOD(T)* $\cap$ *MOD*(ϕ). But when the new information ϕ is in conflict with the current theory T, this no longer works, and we must form a possibly more complex *revision* $T * \phi$. The latter process is quite common, also in ordinary life. When engaged in nonmonotonic reasoning, we may have drawn a temporary conclusion on the basis of just reasoning about minimal models representing some 'normal' scenario. Now something abnormal occurs, and we must accommodate this. Then some of the earlier conclusions no longer hold, and we seek a new equilibrium. Intuitively, revision seems a much more complex process than update, and also, it seems less deterministic. What the new theory will look like depends on what we wish to preserve from the old theory.

Belief revision theory

AGM-style belief revision theory is an abstract postulational account of three basic operations on theories: update, revision, and *contraction* $T - \phi$. The latter removes a proposition from a given theory, and tries to keep the resulting theory as much like the original as possible. The three operations are intertwined:

$$\begin{array}{ll} T * \phi = (T - \neg\phi) + \phi & \text{Levi Identity} \\ T - \phi = T \cap (T * \neg\phi) & \text{Harper Identity} \end{array}$$

Whether these are *all* plausible operations on theories remains an interesting question. But even with just these three, a large literature has sprung up on plausible postulates and concrete mechanisms for performing these operations. Also, the framework has been re-described in various ways, e.g., in terms of Quinean 'entrenchment relations' encoding an agent's willingness to give up one proposition in T rather than another. Similar views of 'structured theories' have come up in computer science when dealing with data bases that may need to be revised in some systematic manner [Ryan, 1992]. The idea that the three operations $\{+, *, -\}$ should yield a unique output may seem strange, as there are usually different options for accommodating conflicting evidence. Accordingly, nondeterministic relational alternatives have been proposed in [Lindström and Rabinowicz, 1992]. At the least, intuitively there should be room for different belief revision *policies*, corresponding to types of behavior: ranging from more 'radical' to more 'conservative' in hanging on to what was there in the old T.

Conditional logic again

Again, there is a remarkable continuity in ideas. The above Ramsey test for conditionals already contained an element of belief revision, as antecedents conflicting one's current stock of beliefs needed to be 'accommodated'. The applicability of the Ramsey test for belief revision has been questioned, but it has become clear that there is nothing wrong with it — pace the 'impossibility results' of Gärdenfors and others (cf. [Leitgeb, 2005]). Indeed, Ramsey-style axioms are a key ingredient in designing modern logics of belief change [Gerbrandy, 1999], without doing any harm in that setting. The connection with conditional logic becomes even clearer in the semantic account of belief revision in Grove [1988]. Viewing a theory as its set of models *MOD(T),* forming a revision $T * \phi$ means passing to the set of T-closest worlds in $MOD(\phi)$, with closeness given by some Lewis-style ordering of theories — or underneath: a similar ordering of worlds. One easily recognizes the key minimality clause underlying the semantics of conditionals. Indeed, conditional statements may be viewed as parts of our current theory, 'pre-encoding' our tendencies to revise our beliefs when confronted with new evidence.

Computer science and AI

Concrete instances of belief revision occur in computer science, with data base maintenance, or belief update in multi-agent systems. Not just conditional logic, but nonmonotonic reasoning in general requires belief revision as a necessary antidote to its 'jumping to conclusions'. Further topics that have been introduced along these lines are the mixture of belief revision and factual update about real changes that have taken place in the world [Katsuno and Mendelzon, 1991].

Learning theory

Belief revision, in its guise of belief adaptation, is actually close to a broader recent stream in the philosophy of science. Perhaps the key activity tied up with theory change is *learning*, whether by individuals or whole communities. Modern learning theory [Osherson *et al.*, 1988; Kelly 1996] describes learning procedures over time, as an account of scientific methods in the face of steadily growing evidence, including surprises contradicting one's current conjecture. In this perspective, update, revision, and contraction are single steps in a larger process, whose temporal structure needs to be brought out explicitly [Kelly, 2002; Hendricks, 2002]. Learning theory is itself a child of recursion theory, and hence it is one more illustration of a computational influence entering philosophy.

Further sources of dynamics

There are many other aspects to update and revision. In particular, another fundamental aspect of the Dynamic Turn is the *social interactive* character of many relevant activities: dialogue, communication, or — for that matter — scientific research. We will look at some of these in §7. To conclude, we return to the simple world of common sense and natural language described in §3. Both update and revision occur all the time in natural discourse. But even this simple setting shows that more is the matter than just automatic *AGM* operations. If I hear some ϕ which contradicts my beliefs so far, then I will need to decide whether to accept the information ϕ, or reject it, or reinterpret it. And a typical way of doing that is to enter a dialogue with ϕ's source, whether a person or Nature. Belief revision usually involves communication and even negotiation. Processes of the latter sort are, again, the subject of §7 below. Another basic feature of revision in common sense settings is the phenomenon of *language adjustment*. The medieval adage was already that "in case of a contradiction, make a distinction". We resolve contradictions either by retracting previous beliefs, or by changing their formulation. New parameters may occur, or new distinctions in the domain of objects. Perhaps you are rich in spirit, while not being rich in material goods. Perhaps Cantor's comprehension axiom creates classes, but not sets. Weinberger [1965] contains an excellent historical and logical survey of basic responses to new information contradicting existing beliefs. In particular, the idea of language change goes beyond standard theories of belief revision. It seems closer to the more dramatic shifts in conceptual frameworks of scientific theories under pressure from recalcitrant facts, as described by Kuhn. Rott [2001b] has a discussion of belief revision in the presence of language change.

From secure foundations to intelligent repair

Belief revision is one more current arena where philosophy meets with logic, language, and computation. Add modern theories of learning to the picture, and we have one of the liveliest interfaces today. In the light of all this, the original emphasis in the foundations of mathematics may have been misguided. Frege wanted absolute security, otherwise "mathematics will come down like a house of cards". Gödel's results show that such security cannot be had, except in tiny corners. But the key phenomenon in understanding rational behaviour does not seem to lie here anyway. Nobody believes that the discovery of a contradiction in a mathematical theory is the end of the world. In fact, every time this has happened (and it does happen), a misunderstanding was exposed, a better and richer theory was found, and a deeper understanding resulted. What seems crucial about us is not the use of infallible methods, but reasoning with whatever means we have, plus an amazing facility for belief revision, i.e., *coping with problems* as they arise. It is this dynamic feature of human rationality which logicians and philosophers should try to understand better. Logic is not some vaccination campaign eradicating all diseases once and for all. It is rather *the immune system of the mind*!

7 DYNAMIC LOGIC, COMMUNICATION, ACTION, AND GAMES

A logical consequence is often described as involving a conclusion which 'adds no new information to the premises'. Indeed, logic is deeply tied up with extracting and modifying information, as well as knowledge and other epistemic attitudes arising in these informational processes. Defining a precise notion of 'information' serving all logical purposes is no trivial task (cf. [Adriaans and van Benthem, 2005], *Handbook of the Philosophy of Information*). There are several different strands in the logical literature (cf. [Carnap, 1947; Barwise and Seligman, 1995]), whose connections are not always obvious [van Benthem, 2005b]). The story line in this section follows one of these, viz. information as a range of options for the actual world that are still compatible with what agents know. This was the original idea in Carnap's *Meaning and Necessity*, and it became central later on in the classical work on epistemic logic [Hintikka, 1962]. Epistemic logic is an interesting place to start a story line, since many people consider it a typical case of a dead-end as an interface between logic and philosophy. This is as far from the truth as it gets, and we give a short summary of [van Benthem, 2005a].

Statics: epistemic logic and information states

Epistemic logic started as an account of the philosophical notion of 'knowledge', for which we have famous accounts by Plato or Descartes. The second half of the 20th century added a host of interesting alternatives, including contributions by Gettier [1963], Hintikka [1962], Dretske [1981], and Nozick [2001]. Hintikka's main idea was this. An agent *knows* those propositions which are true in all situations compatible with what she knows about the actual world:

$$\mathbf{M}, s \models K_i\phi \text{ iff for all } t \sim_i S : \mathbf{M}, t \models \phi.$$

New information will tend to shrink this range, perhaps until the actual world is all that is left open to our scrutiny. There are similar accounts of *belief*, and other epistemic attitudes. The power of this formalism shows in simple scenarios. Consider the simplest case of genuine information flow between agents:

> **Q** asks a question "*P*?", and **A** gives a true answer "Yes".

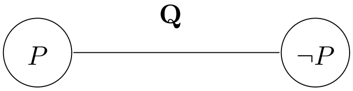
If this is all honest Gricean communication, then the question itself indicates that the **Q** does not know the answer (base-level factual knowledge), but also, that she thinks **A** might know (knowledge about other people's information). The answer conveys that that Answerer knows that P, and its public announcement in the group $\{\mathbf{Q}, \mathbf{A}\}$ makes sure that **Q** now also knows that P, that both agents know this about each other, and so on. In terms of Lewis [1969], following the episode, the agents have achieved *common knowledge* of P. Epistemic logic as it stands can describe all relevant epistemic properties at the three stages of this process: before, during, and after. And of course, it can do a lot more, witness the further literature on this subject [Hendricks, 2005].

Now, in this area, one of the most spectacular 'crossings' took place. First, in the mid-1970s, epistemic logic emerged independently in game theory [Aumann 1976; Osborne and Rubinstein, 1994], as economists developed epistemic scenarios for justifying the usual Nash equilibria of strategies. The idea was to 'lock' players 'into' equilibria in terms of mutual expectations (via knowledge or belief) concerning each other's rationality. Also, by the 1980s, epistemic logic crossed over to computer science [Fagin *et al.*, 1995]. Here it was the evident power of human-style models for understanding protocols in distributed systems that started a vast literature at the interface of computer science, philosophy, and game theory. It is about agents who (want to) know things, and communicate accordingly. The TARK conferences (*Theoretical Aspects of Rationality and Knowledge*) have brought together these three communities since 1983, and the new interdisciplinary journal *Knowledge, Rationality, and Action* show that the contact is still very much alive.

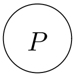
In this section, we just sketch one particular strand in this story, tying it up with the logical dynamics that emerged in §6. Knowledge of a proposition comes about by certain *actions*, be it an inference, asking or answering a question, or performing an observation of some event in general. As before, the static structure of the knowledge and the dynamic nature of those actions are intertwined. Thus, to really get at what happened in our question-answer scenario, we must 'dynamify' standard epistemic logic!

Updating information

Speech act theory in the philosophy of language tells us that linguistic communication triggers systematic changes in information states of language users. For instance, the above question-answer episode might start from the following simple epistemic model, where the line between the two alternative worlds indicates my uncertainty of the questioner **Q**:



Since you have no uncertainty lines in either world, you know whether P throughout, and I actually know that you know, as it is true in both of my alternatives. Now, your affirmative answer to my question *updates* this model, ruling out the $\neg P$-world, to obtain:



Here both you and I know that P, and we know this about each other, and so on to arbitrary further mutual stages: P has become common knowledge between the two of us.

Notice what happened here in more general terms: epistemic logic now has to deal with dynamic actions, while there is also an irreducible multi-agent 'social' character to the setting.

Dynamic-epistemic logic

Both social and private epistemic actions can be described as first-class citizens in dynamified epistemic logics that take a leaf from computer science. Our epistemic actions that eliminate worlds are really model-changers, and this process involves constant changes in truth value for relevant assertions. Similar phenomena occur with computer programs, recipes, or any sort of structured action. Such phenomena have been studied in *dynamic logics* (Hoare, Dijkstra, Pratt), whose languages involve mixed assertions with action expressions and propositions. In particular, a dynamic modality $[a]\phi$ says that, after every successful execution of action or program a from the current state of the world, or our device, ϕ is the case. In this style, we would get

$[A!]\phi$ after public announcement of A, formula ϕ holds.

This mixed proposition/action language can express epistemic effects such as

$[A!]K_jB$ after public announcement of A, agent j knows that B
$[A!]C_GA$ likewise, A has become common knowledge in the group of agents G.

There are complete and decidable dynamic-epistemic logics of this kind, allowing us to reason systematically about the epistemic effects of assertions. Here is an example of a valid 'update law' — which is as central to dynamic-epistemic logic as an introspection principle would be to static epistemic logic:

$$[A!]K_i\phi \leftrightarrow (A \to K_i[A!]\phi) \qquad \#$$

Note the complex intellectual history concentrated in formulas like this. Speech acts come from the philosophy of language and linguistics, epistemic logic from epistemology, and their combination in a dynamic logic comes from computer science. This harmonious encounter is a perfect illustration of the story lines of this paper, and the axiomatic calculus suggests new contacts between these fields. E.g., philosophers of action will note that the axiom # expresses an interchange between knowledge that we have beforehand about an action and knowledge that we have after the action. This is a nontrivial feature, which a game-theorist would call *Perfect Recall*: public announcements are epistemically transparent. By contrast, e.g., an action like drinking would not satisfy this commutation axiom. I know that drinking will make me boring — but after drinking, alas, I do not know that I am boring.

Thus, the dynamic turn in epistemic logic greatly enriches the agenda of interesting issues for philosophers, allowing them to move beyond worries about Omniscience, Introspection, and the other usual suspects.

More complex communication over time

Dynamic-epistemic logics do not just deal with single assertions. They can also handle complex conversational instructions involving typical sequential program structures, such as "First say this, then that", "Say 'Thanks' if you are grateful, and 'No thanks' otherwise", "Keep saying nice things until the dean grants your promotion", or even a parallel program structure like "Answer all together".

Finally, current systems can deal with much more complex, and realistic, communicative actions which involve hiding of information, or partial observation. This includes whispering in lecture rooms, looking at your hand in a card game, or many further informational moves in games that people play. Some basic sources are [Baltag *et al.*, 1998; van Ditmarsch *et al.*, 2005; van Benthem 2002b; 2005e]. In particular, dynamic-epistemic logic can deal with misleading actions (hiding, lying, cheating) as well as truthful ones — leading to a classification of epistemic action as to expressive power and thresholds of complexity. Also, the framework links up with *belief revision*, *learning*, *probabilistic update*, as well as infinite processes and evolution over time.

Philosophy again

Update logic may seem new-fangled, but it connects up with well-established issues in philosophy. Van Benthem [2004] discusses the age-old Moore-sentences of the possibly true but pragmatically infelicitous form "p, but I do not know it". These turns up in a surprising new setting in epistemic dynamics. It seems evident at first that public announcements of true propositions. A lead to common knowledge of A. Indeed, this seems to be the purpose of public announcements, as described by speech act theorists. But now consider an assertion like

> "you don't know if p, but p is true".

This can be perfectly true, and yet its announcement makes p common knowledge, thereby invalidating the first conjunct. Thus, this proposition is 'self-refuting'. This failure may seem like a trick, but announcements of uncertainty leading eventually to knowledge occur quite often in famous puzzles, such as The Wise Men or the Muddy Children. True but self-refuting assertions are also at the heart of current discussions of the Fitch Paradox [Brogaard and Salerno, 2002] threatening the Verificationist Thesis that "everything which is true can be known".

More generally, dynamic-epistemic logics shift the traditional boundary between semantics and pragmatics in interesting ways. This demarcation issue is on the agenda in both linguistics and philosophy of language these days. Another interface issue concerns the topic of theory change in §7. The internal information flow of a scientific community can be analyzed now, not just the public theories produced by that community. What are the systematic relationships between the short-term multi-agent dynamics of scientific debate and the long-term dynamics of theory change?

Games

In its current manifestations, logical dynamics is making one more move toward *game theory.* Both knowledge and communication really make sense only against the background of goal-directed interactive social processes between rational agents. Van Benthem [1999-present; 2005d] provide an extensive panorama of how games interface with logic. This includes 'logic games' for argumentation [Lorenz and Lorenzen, 1978] and semantic evaluation [Hintikka, 1973], showing how these core tasks of logic may be cast as interactive processes between different agents. But the interface also involves modern 'game logics' that have been used to throw light on the rationality of players in general games [Aumann, 1976; Stalnaker, 1999] and earlier publications, [Pauly, 2001; de Bruin, 2004]. In these game logics, many of our earlier systems come together: dynamic logic, epistemic logic, but also logics of counterfactuals, and belief revision. Games started in philosophy with Wittgenstein's language games, and they were advocated vigorously by Hintikka through his 'game-theoretic semantics' and '*IF*-logic' [Hintikka and Sandu, 1997]. But it seems we have only seen the beginning of their true impact in a broader sense (cf. [Hendricks, 2005] on interactive epistemology).

Thus again, we see how the current philosophical agenda could be enriched by fully absorbing the fundamental nature of multi-agent information and interaction.

8 FURTHER THEMES

The themes in this paper are just a selection from a much longer list, including several survey papers written by this author in other volumes. Other story lines through the last decades linking logic, philosophy and computer science for which the material lies at hand include the following. The references given are just pointers — not a complete portal to these areas:

(a) *structure of scientific theories* (cf. the list of 'Semantic Parallels' in van Benthem [1989], which is still only partially explored)

(b) *knowledge and belief* [van Benthem, 2005a])

(c) *information and computation* [van Benthem, 2005c]

(d) *probability*

(e) *temporal logic* [van Benthem, 1995]

(f) *deontic logic, norms and obligations* [van Hees and Pauly, 2004], as well as the Proceedings of the *DEON* conferences, operating since the late 1980s at the interface of computer science and logic)

(g) *philosophy of action* [Belnap *et al.*, 2001]

(h) *infinite processes and dynamical systems* [Skyrms, 2004]

(i) *optimality and evolutionary games* [van Rooy, 2004].

Thus, the number of strands connecting logic and philosophy is large, and the fabric of the interface is hard to tear accordingly.

9 CONCLUSION: LOGIC AND PHILOSOPHY ONCE MORE

The upshot of this paper is simply this. The history of interfaces between logic and philosophy is rich and varied, especially when we describe it through themes, rather than formal languages or systems. Our story lines are just sketchy first illustrations, but they demonstrate the broad spirit. Before we can have any considered opinion of the current state of affairs between logic and philosophy, we need such a rich view of their past and present interactions. The resulting picture may not be quite like received opinions as to what are the most crucial topics — but it seems definitely much richer than what would be found in any standard text book. It may even help get rid of some internal research programs in 'philosophical logic' that seem to have run their course.

As for more practical conclusions, past meetings between logic and philosophy have been successful, and there seems every reason for pursuing these contacts. Moreover, these contacts are often enhanced through a third party, viz. computer science, which acts as a laboratory for continuing philosophy 'by other means'. Lively interdisciplinary conferences like TIME, CONTEXT, or TARK lead the way in this menage à trois.

Still, we should freely admit that we have not provided any clear demarcation of 'logic' in our present perspective. What we have in mind is broader than many current definitions; but still, not anything goes. If pressed right now, we would say that 'logic' is better described as a certain rigorous style of working than any very precisely demarcated subject matter. And in any case, the spirit of this essay is that definitions should follow realities, rather than the other way around.

But perhaps most of all, we have written on purpose in a nonconventional style. Much intellectual history seems to make what happened Platonic, heroic, inevitable, and thus: hard to identify with. But our founding fathers were just ordinary people like us, with all sorts of options open to them, including chance encounters, and outside events — and mathematical logic or philosophical logic, though shining with the halo of time, are still open to significant changes, both in terms of interpretation and in terms of real change. Just read Stelzner [1996] on how the optical firm of Carl-Zeiss Jena sponsored Frege's tenure case (with the physicist Ernst Abbe as a guardian angel), and your view of the birth of modern logic, the genesis of *Begriffsschrift*, and its famous metaphor of Microscope versus Eye, will never be the same again!

ACKNOWLEDGEMENTS

I would like to thank Hannes Leitgeb, Gabriel Sandu, Frank Veltman, and Dag Westerståhl for their useful and congenial comments.

BIBLIOGRAPHY

[Adriaans and van Benthem, 2005] P. Adriaans & J. van Benthem, eds., 2005, *Handbook of the Philosophy of Information*, Elsevier, Amsterdam, to appear.

[Alchourron *et al*, 1985] C. Alchourron, P. Gärdenfors & D. Makinson, 1985, On the Logic of Theory Change: Partial Meet Contraction and Revision Functions, *Journal of Symbolic Logic* 50, 510–530.

[Aliseda-Llera, 1997] A. Aliseda-Llera, 1997, *Seeking Explanations: Abduction in Logic, Philosophy, and Artificial Intelligence*, Dissertation, Department of Philosophy, Stanford University & ILLC, University of Amsterdam.

[Aumann, 1976] R. Aumann, 1976, Agreeing to Disagree, *Annals of Statistics* 4, 1236–1239.

[d'Avila Garcez *et al.*, 2002] A. d'Avila Garcez, K. Broda & D. Gabbay, 2002, *Neural-Symbolic Learning Systems: Foundations and Applications*, Springer, London.

[Baltag *et al.*, 1998] A. Baltag, L. Moss & S. Solecki, 1998, The Logic of Public Announcements, Common Knowledge and Private Suspicions, *Proceedings TARK 1998*, 43–56, Morgan Kaufmann Publishers, Los Altos.

[Barth, 1974] E. Barth, 1974, *The Logic of the Articles in Traditional Philosophy*, Reidel, Dordrecht.

[Barwise and Perry, 1983] J. Barwise & J. Perry, 1983, *Situations and Attitudes*, The MIT Press, Cambridge (Mass.).

[Belnap *et al.*, 2001] N. Belnap, M. Perloff & M. Xu, 2001, *Facing the Future*, Oxford University Press, Oxford.

[van Benthem, 1981] J. van Benthem, 1981, Historische Vergissingen? Kanttekeningen bij de Fregeaanse Revolutie (*Historical Errors? Marginal comments on the Fregean Revolution in Logic*), *Kennis en Methode* 5:2, 94–116.

[van Benthem, 1982] J. van Benthem, 1982, The Logical Study of Science, *Synthese* 51, 431–472.

[van Benthem, 1985] J. van Benthem, 1985, The Variety of Consequence, According to Bolzano, *Studia Logica* 44:4, 389-403.

[van Benthem, 1986] J. van Benthem, 1986, *Essays in Logical Semantics,* Reidel, Dordecht.

[van Benthem, 1987] J. van Benthem, 1987, Meaning: Interpretation and Inference, *Synthese* 73, 451-470.

[van Benthem, 1989] J. van Benthem, 1989, Semantic Parallels in Natural Language and Computation, in H-D Ebbinghaus et al., eds., *Logic Colloquium. Granada 1987*, North-Holland, Amsterdam, 331-375.

[van Benthem, 1991] J. van Benthem, 1991, *Language in Action*, North-Holland, Amsterdam. Reprinted 1995, The MIT Press, Cambridge (Mass.).

[van Benthem, 1995] J. van Benthem, 1995, Temporal Logic, in D. Gabbay, C. Hoggar & J. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, Oxford University Press, 241–350.

[van Benthem, 1996a] J. van Benthem, 1996A, *Exploring Logical Dynamics*, CSLI Publications, Stanford.

[van Benthem, 1996b] J. van Benthem, 1996B, Inference, Methodology and Semantics, in P. Bystrov & V. Sadofsky, eds., *Philosophical Logic and Logical Philosophy, Essays in Honour of Vladimir Smirnov,* Kluwer Academic Publishers, Dordrecht, 63–82.

[van Benthem, 1999a] J. van Benthem, 1999A, Modal Foundations for Predicate Logic, in E. Orlowska, ed., *Logic at Work*, To the Memory of Elena Rasiowa, Physica Verlag, Heidelberg, 39–54.

[van Benthem, 1999b] J. van Benthem, 1999B, Wider Still and Wider: resetting the bounds of logic, in A. Varzi, ed., *The European Review of Philosophy*, CSLI Publications, Stanford, 21–44.

[van Benthem, 1999–present] J. van Benthem, 1999–present, *Logic in Games*, Electronic lecture notes, ILLC University of Amsterdam & Department of Philosophy, Stanford University. `http://staff.science.uva.nl/~johan/Teaching`

[van Benthem, 2002a] J. van Benthem, 2002A, Mathematical Logic and Natural Language, in B. Löwe, W. Malzkorn, & T. Rasch, eds., *Foundations of the Formal Sciences II: Applications of Mathematical logic in Philosophy and Linguistics*, Kluwer, Dordrecht, 25–38.

[van Benthem, 2002b] J. van Benthem, 2002B, One is a Lonely Number, ILLC Report 2002-07, University of Amsterdam. To appear in P. Koepke & W. Pohlers, eds., *Colloquium Logicum*, ASL Publications, New Providence.

[van Benthem, 2003a] J. van Benthem, 2003A, Fifty Years: Changes and Constants in Logic, in V. Hendricks & J. Malinowski, eds., *Trends in Logic, 50 Years of Studia Logica*, Kluwer, Dordrecht, 35–56.

[van Benthem, 2003b] J. van Benthem, 2003B, Is there still Logic in Bolzano's Key?, in E. Morscher, ed., *Bernard Bolzanos Leistungen in Logik, Mathematik und Physik. Beiträge zur Bolzano-Forschung,* Bd.16, Academia Verlag, Sankt Augustin 2003, pp.11-34.

[van Benthem, 2003c] J. van Benthem, 2003C, Logic and the Dynamics of Information, in L. Floridi, ed., *Minds and Machines* 13:4, 503–519.

[van Benthem, 2003d] J. van Benthem, 2003D, Logic in Philosophy, electronic course notes, `http://staff.science.uva.nl/~johan/Teaching/` , Stanford University.

[van Benthem, 2003e] J. van Benthem, 2003E, Structural Properties of Dynamic Reasoning. in J. Peregrin, ed., *Meaning: the Dynamic Turn*, Elsevier, Amsterdam, 2003, 15-31.

[van Benthem, 2003f] J. van Benthem, 2003F, The Categorial Fine-Structure of Natural Language, Tech Report PP-2003-20, ILLC Amsterdam. In C. Casadio, P.J. Scott, R.A.G. Seely, eds., *Language and Grammar: Studies in Mathematical Linguistics and Natural Language,* CSLI Publications Stanford (2005), 3–29.

[van Benthem, 2004] J. van Benthem, 2004, What One May Come to Know, *Analysis* 64 (282), 95–105.

[van Benthem, 2005a] J. van Benthem, 2005A, Epistemic Logic and Epistemology. The State of their Affairs, to appear in V. Hendricks, ed., special issue of *Philosophical Studies* on Bridges between Formal and Mainstream Epistemology.

[van Benthem, 2005b] J. van Benthem, 2005B, Information as Correlation and Information as Range, Working Paper, ILLC, University of Amsterdam. To appear in L. Moss, ed., *Memorial Volume for Jon Barwise.*

[van Benthem, 2005c] J. van Benthem, 2005C, Information Is What Information Does, draft editorial, *Handbook of Philosophy of Information*, ILLC University of Amsterdam.

[van Benthem, 2005d] J. van Benthem, 2005D, Open Problems in Logic and Games. In *We Will Show Them! Essays in Honour of Dov Gabbay, Volume 1*, S. Artemov, H. Barringer, A. d'Avila Garcez, L. Lamb and J. Woods, eds. pp. 229–264. College Publications, 2005.

[van Benthem, 2005e] J. van Benthem, 2005E, Open Problems in Logical Dynamics, ILLC Preprint:, Institute for Logic, Language and Information, University of Amsterdam, DARE electronic archive 148382. To appear in D. Gabbay, S. Goncharov & M. Zakharyashev, eds., *Mathematical Problems from Applied Logic*, Russian Academy of Sciences, Novosibirsk & Springer, New York.

[van Benthem and ter Meulen, 1997] J. van Benthem & A. ter Meulen, eds., 1997, *Handbook of Logic and Language*, Elsevier, Amsterdam.

[Beth, 1949] E. Beth, 1949, Towards an Up-to-Date Philosophy of the Natural Sciences, *Methodos1*, 178–185.

[Blanshard, 1964] B. Blanshard, 1964, *Reason and Analysis*, Open Court Publishers, La Salle (Ill.).

[Bolzano, 1937] B. Bolzano, 1837, *Wissenschaftslehre*, Buchhandlung Seidel, Sulzbach. Translated as *Theory of Science* by R. George, University of California Press, Berkeley & Los Angeles, 1972.

[Brogaard and Salerno, 2002] B. Brogaard & J. Salerno, 2002, Fitch's Paradox of Knowability, Stanford Electronic Encyclopedia of Philosophy, `http://plato.stanford.edu/entries/fitch-paradox/`.

[de Bruin, 2004] B. de Bruin, 2004, *Explaining Games*, Dissertation 2004-03, ILLC Amsterdam.

[Carnap, 1947] R. Carnap, 1947, *Meaning and Necessity*, University of Chicago Press, Chicago.

[CONTEXT, 2005] CONTEXT, 2005, 5th International and Interdisciplinary Conference on Modeling and Using Context, July 5-8, Paris, France, `www.context-05.org`.

[Davidson, 1967] D. Davidson, 1967, Truth and Meaning, *Synthese* 17, 304–323.

[Davidson and Harman, 1972] D. Davidson & G. Harman, eds., 1972, *Semantics of Natural Language*, Reidel, Dordrecht.

[van Ditmarsch *et al.*, 2005] H. van Ditmarsch, W. van der Hoek & B. Kooi, 2005, *Dynamic Epistemic Logic*, Kluwer Academic Publishers, Dordrecht, to appear.

[Dosens and Schröeder-Heister, 1994] K. Dosen & P. Schröder-Heister, eds., 1994, *Substructural Logics*, Clarendon Press, Oxford.

[Dretske, 1981] F. Dretske, 1981, *Knowledge and the Flow of Information*, Chicago University Press, Chicago .

[Dummett, 1973] M. Dummett, 1973, *Frege, the Philosophy of Language*, Duckworth, London.

[Dutilh-Novaes, 2003] C. Dutilh-Novaes, 2003, Medieval Obligationes as Logical Games of Consistency, Philosophical Institute, University of Leiden.

[Englebretsen, 1981] G. Englebretsen, 1981, *Three Logicians*, van Gorcum, Assen.

[Etchemendy, 1990] J. Etchemendy, 1990, *The Concept of Logical Consequence*, Harvard University Press, Cambridge (Mass.).

[Fagin *et al.*, 1995] R. Fagin, J. Halpern, Y. Moses & M. Vardi, 1995, *Reasoning about Knowledge*, The MIT Press, Cambridge (Mass.).

[Field, 1972] H. Field, 1972, Tarski's Theory of Truth, *Journal of Philosophical Logic* 69, 347-375.

[Flach and Kakas, 2000] P. Flach & A. Kakas, eds., 2000, *Abduction and Induction. Essays on their Relation and Integration*, Kluwer Academic Publishers, Dordrecht.

[Floridi, 2004] L. Floridi, ed., 2004, *The Blackwell Guide to the Philosophy of Computing and Information*, Blackwell, Oxford – New York.

[Gabbay and Guenthner, 1983–1988] D. Gabbay & F. Guenthner, eds., 1983–1988, *Handbook of Philosophical Logic*, Reidel, Dordrecht. Second revised and enlarged edition starting from 1996, Kluwer Academic Publishers, Dordrecht.

[GAMUT, 1991] GAMUT, 1991, *Logic, Language and Meaning*, Chicago University Press.

[Gärdenfors, 1987] P. Gärdenfors, 1987, *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, The MIT Press, Cambridge (Mass.).

[Gärdenfors and Rott, 1995] P. Gärdenfors & H. Rott, 1995, Belief Revision, in *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, ed. by D. M. Gabbay, C. J. Hogger, and J. A. Robinson, Oxford University Press, Oxford.

[Geach, 1972] P. Geach, 1972, *Logic Matters*, Blackwell. Oxford.

[Gerbrandy, 1999] J. Gerbrandy, 1999, *Bisimulations on Planet Kripke*, dissertation DS-1999-01, Institute for Logic, Language and Computation, University of Amsterdam.

[Gettier, 1963] E. Gettier, 1963, Is Justified True Belief Knowledge?, *Analysis* 23, 121–123.

[Geurts and van der Slik, 2004] B. Geurts & F. van der Slik, 2004, Monotonicity and Processing Load, *Journal of Semantics* 22, 97-117.

[Gillies, 2004] A. Gillies, 2004, New Foundations for Epistemic Change, *Synthese* 138, 1–48.

[Grice, 1975] H. Grice, 1975, Logic and Conversation, in P. Cole & J. Morgan, eds., *Syntax and Semantics III*, Academic Press, New York, 41–58.

[Groenendijk and Stokhof, 1991] J. Groenendijk & M. Stokhof, 1991, Dynamic Predicate Logic, *Linguistics & Philosophy* 14, 39–100. Also in *The Philosopher's Annual*, 1992.

[Grove, 1988] A. Grove, Two Modelings for Theory Change, *Journal of Philosophical Logic*, 157-170, 1988.

[van Hees and Pauly, 2004] M. van Hees & M. Pauly, 2004, Logical Constraints on Aggregating Judgments, Department of Philosophy, University of Groningen & INRIA Toulouse. To appear in the *Journal of Philosophical Logic*.

[Hempel, 1965] P. Hempel, 1965, *Aspects of Explanation and Other Essays in the Philosophy of Science*, The Free Press, New York.

[Hendricks, 2002] V. Hendricks, 2002, Active Agents, PHILOG Newsletter, Roskilde. In J. van Benthem & R. van Rooy, eds., special issue on Information Theories, *Journal of Logic, Language and Information* 12:4, 469–495.

[Hendricks, 2005] V. Hendricks, 2005, *Mainstream and Formal Epistemology*, Cambridge University Press, New York.

[Hintikka, 1962] J. Hintikka, 1962, *Knowledge and Belief*, Cornell University Press, Ithaca. Reprint 2005, V. Hendricks & J. Symons, eds., King's College Publications, London.

[intikka, 1973] J. Hintikka, 1973, *Logic, Language Games and Information*, Oxford University Press, Oxford.

[Hintikka and Sandu, 1997] J. Hintikka & G. Sandu, 1997, 'Game-Theoretical Semantics, in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier, Amsterdam, 361–410.

[Hobbs *et al.*, 1990] J. Hobbs, M. Stickel, D. Appelt & P. Martin, 1990, *Interpretation as Abduction*, SRI Technical Note 499, SRI International, Menlo Park, California.

[Hodges, 2001] W. Hodges, 2001, Formal Features of Compositionality, *Journal of Logic, Language and Information* 10, 7–28.

[Hodges, 2005] W. Hodges, 2005, Opening Two Windows: Arabic Logic and Cognitive Science, Department of Mathematics, Queen Mary and William's College, London. To appear in D. Gabbay, S. Gontcharov & M. Zakharyashev, eds., *Logics for the XXIst Century: Mathematical Problems From Applied Logic*, Russian Academy of Sciences, Novosibirsk & Kluwer/Plenum Publishers, New York.

[Janssen, 1981] Th. Janssen, 1981, *Foundations and Applications of Montague Grammar*, Dissertation, Mathematical Institute, University of Amsterdam.

[Jacquette, 2002] D. Jacquette, ed., 2002, *A Companion to Philosophical Logic*, Blackwell, Oxford.

[Kamp and Reyle, 1993] H. Kamp & U. Reyle, 1993, *From Discourse to Logic*, Kluwer, Dordrecht.

[Katsuna and Mendelzon, 1991] H. Katsuno & A. Mendelzon, 1991, "Propositional Knowledge Base Revision and Minimal Change, *Artificial Intelligence* 52, 263–294.

[Keenan and Westerståhl, 1997] E. Keenan & D. Westerståhl, 1997, Quantifiers, in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier, Amsterdam, 837–893.

[Kelly, 1996] K. Kelly, 1996, *The Logic of Reliable Enquiry*, Oxford University Press, Oxford.

[Kelly, 2002] K. Kelly, 2002, Knowledge as Reliably Inferred Stable True Belief, Department of Philosophy, Carnegie-Mellon University, Pittsburgh.

[Kripke, 1975] S. Kripke, 1975, Outline of a Theory of Truth, *Journal of Philosophy* 72, 690–716.

[Kuipers, 2000] Th. Kuipers, 2000, *From Instrumentalism to Constructive Realism*, Kluwer Academic Publishers, Dordrecht.

[Kuhn, 1962] Th. Kuhn, 1962, *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago.

[Leitgeb, 2004a] H. Leitgeb, 2004A, Carnap's Logischer Aufbau Revisited, Department of Philosophy, University of Salzburg.

[Letigeb, 2004b] H. Leitgeb, 2004B, *Inference on the Low Level. An Investigation into Deduction, Nonmonotonic Reasoning, and the Philosophy of Cognition*, Springer, Berlin.

[Leitgeb, 2005] H. Leitgeb, 2005, Gärdenfors' Impossibility Argument Defused, working paper, Department of Philosophy, Stanford University & Fachbereich Philosophie, Universität Salzburg.

[Lewis, 1969] D. Lewis, 1969, *Convention*, Harvard University Press, Cambridge (Mass.).

[Lewis, 1973] D. Lewis, 1973, *Counterfactuals*, Blackwell, Oxford.

[Lindström and Rabinowicz, 1992] S. Lindström & W. Rabinowicz, 1992, Belief Revision, Epistemic Conditionals and the Ramsey Test, *Synthese* 91,195-237.

[Lorenz and Lorenzen, 1978] K. Lorenz & P. Lorenzen, 1978, *Dialogische Logik*, Wissenschaftliche Buchgesellschaft, Darmstadt.

[Mates, 1953] B. Mates, 1953, *Stoic Logic*, University of California Publications, Berkeley & Los Angeles.

[McCarthy, 1980] J. McCarthy, 1980, Circumscription – A Form of Nonmonotonic Reasoning, *Artificial Intelligence* 13, 27–39.

[McCarthy, 2005] J. McCarthy, 2005, The Philosophy of AI and the AI of Philosophy, Department of Computer Science, Stanford University. To appear in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier, Amsterdam.

[Montague, 1974] R. Montague, 1974, *Formal Philosophy, Selected Papers of Richard Montague*, R. Thomason, ed., Yale University Press, New Haven.

[Nozick, 1981] R. Nozick, 1981, *Philosophical Explanations*, Harvard University Press, Cambridge (Mass.).

[Osherson *et al.*, 1988] D. Osherson, M. Stob & S. Weinstein, 1988, *Systems that Learn*, The MIT Press, Cambridge (Mass.).

[Parikh, 2001] P. Parikh, 2001, *The Use of Language*, CSLI Publications, Stanford.

[Pauly, 2001] M. Pauly, 2001, *Logic for Social Software*, dissertation DS-2001-10, Institute for Logic, Language and Computation, University of Amsterdam.

[Pearce and Rantala, 1984] D. Pearce & V. Rantala, 1984, A Logical Study of the Correspondence Relation, *Journal of Philosophical Logic* 13, 47–84.

[Pereira and Warren, 1983] F. Pereira & D. Warren, 1983, Parsing as Deduction, *Proceedings of the 21st Meeting of the Association for Computational Linguistics*, 137–144.

[Peters and Westerståhal, to appear] S. Peters & D. Westerståhl, to appear, *Quantifiers in Language and Logic*, Philosophical Institute, University of Gothenburg & Department of Linguistics, Stanford University. To appear with Oxford University Press.

[Przelecki, 1969] M. Przelecki, 1969, *The Logic of Empirical Theories*, Routledge and Kegan Paul, London.

[Putnum, 1975] H. Putnam, 1975, *Mind, Language, and Reality*, Cambridge University Press, Cambridge.

[ramsey, 1931] F. Ramsey, 1931, *The Foundations of Mathematics and Other Essays*, Kegan Paul, London.

[Reiter, 2001] R. Reiter, 2001, *Knowledge in Action. Logical Foundations for Specifying and Implementing Dynamical Systems*, The MIT Press, Cambridge (Mass.).

[de Roever *et al.*, 2001] W-P de Roever, F. de Boer, U. Hannemann, J. Hooman, Y. Lakhnech, M. Poel & J. Zwiers, 2001, *Concurrency Verification: Introduction to Compositional and Noncompositional Methods*, Cambridge University Press, Cambridge.

[van Rooy, 2004] R. van Rooy, 2004, Signalling Games Select Horn Strategies, *Linguistics and Philosophy* 27, 493-527.

[van Rooy and Schultz, 2004] R. van Rooy & K. Schultz, 2004, Pragmatic Meaning and Non-Monotonic Reasoning: The Case of Exhaustive Interpretation, Working Paper, ILLC University of Amsterdam. To appear in *Linguistics and Philosophy.*

[Rott, 2001] H. Rott, 2001A, *Change, Choice, and Inference*, Clarendon Press, Oxford.

[rott, 2001b] H. Rott, 2001B, Theoretical Concepts in Flux: Conceptual Knowledge and Theory Change, in R. Eckardt, K. von Heusinger & Ch. Schwarze, eds., *Words in Time – Diachronic Semantics from Different Points of View*, FB Sprachwissenschaft der Universität Konstanz, 279–301.

[Ryan, 1992] M. Ryan, 1992, *Ordered Presentations of Theories: Default Reasoning and Belief Revision*, Ph.D. thesis, Department of Computing, Imperial College.

[Sanchez Valencia, 1991] V. Sanchez Valencia, 1991, *Studies in Categorial Grammar and Natural Logic*, Dissertation, ILLC, University of Amsterdam. To appear with MIT Press.

[Shanahan, 1997] M. Shanahan, 1997, *Solving the Frame Problem*, The MIT Press, Cambridge (Mass.).

[Shoham, 1988] Y. Shoham, 1988, *Reasoning about Change. Time and Causation from the Standpoint of AI*, The MIT Press, Cambridge (Mass.).

[Skyrms, 2004] B. Skyrms, 2004, *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press, Cambridge.

[Smolensky and Legendre, 2005] P. Smolensky & G. Legendre, 2005, *The Harmonic Mind. From Neural Computation to Optimality-Theoretic Grammar*, The MIT Press, Cambridge (Mass.).

[Sommers, 1982] F. Sommers, 1982, *The Logic of Natural Language*, Clarendon, Oxford.

[Stalnaker, 1999] R. Stalnaker, 1999, Extensive and Strategic Form: Games and Models for Games, *Research in Economics* 53:2, 93–291.

[Stelzner, 1996] W. Stelzner, 1996, *Gottlob Frege. Jena und die Geburt der Modernen Logik*, Verein zur Regionalforderung von Forschung, Innovation und Technologie, Jena.

[Suppe, 1977] F. Suppe, ed., 1977, *The Structure of Scientific Theories*, University of Illinois Press, Urbana.

[Tarski, 1956] A. Tarski, 1956, *Logic, Semantics, Metamathematics*, Oxford University Press, Oxford.

[van Ulsen, 2000] P. van Ulsen, 2000, *Evert Willem Beth als Logicus*, Dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

[Veltman, 1996] F. Veltman, 1996, Defaults in Update Semantics, *Journal of Philosophical Logic* 25, 221–261. Also appeared in *The Philosopher's Annual*, 1997.

[Wason and Johnson-Laird, 1972] P. Wason & P. Johnson-Laird, 1972, *Psychology of Reasoning. Structure and Content*, Batsford, London.

[Weinberger, 1965] O. Weinberger, 1965, Der Relativisierungsgrundsatz und der Reduktionsgrundsatz – Zwei Prinzipien des Dialektischen Denkens, Czechoslovakian Academy of Sciences, Prague.

[Wundt, 1880–3] W. Wundt, 1880/3, *Logik. Eine Untersuchung der Prinzipien der Erkenntniss und der Methoden Wissenschaftlicher Forschung*, Ferdinand Enke, Stuttgart.

[Zalta, 1993] E. Zalta, 1993, A Philosophical Conception of Propositional Modal Logic, *Philosophical Topics* 21:2, 263–281.

INFORMAL LOGIC AND THE CONCEPT OF ARGUMENT

David Hitchcock

According to its namers, informal logic "is best understood as the normative study of argument. It is the area of logic which seeks to develop standards, criteria and procedures for the interpretation, evaluation and construction of arguments and argumentation used in natural language." [Blair and Johnson, 1987, 148]; similarly, [Johnson and Blair, 2000, 94] The name "informal logic" is somewhat unfortunate. For those who use "logical" as a synonym of "formal", it is an oxymoron. In any case, the research programme of informal logic does not preclude the use of formal methods or appeal to formal logics. Its distinctiveness consists in its consideration of a set of questions that are not addressed in the specialist journals of formal logic, such as the *Journal of Symbolic Logic* and the *Notre Dame Journal of Formal Logic*, or in such histories of formal logic as that by William and Martha Kneale [1962]. It might in fact better be called "theory of argument". Its questions have however traditionally been regarded as part of logic, broadly conceived. The name can thus be taken to refer to that part of logic as traditionally conceived that is not covered by contemporary formal logic.

Johnson and Blair [2000, 99–100] group the questions investigated by informal logic under the headings of argument identification, argument analysis, argument evaluation, and argument criticism. In accordance with the definition quoted in the opening sentence of this chapter, one should add argument construction as a fifth heading. Because of limitations of space and time, this chapter deals intensively with just one question within informal logic, the question: What is an argument? A postscript lists other questions about arguments investigated within informal logics, and refers to the relevant literature.

TECHNICAL AND EVERYDAY SENSES OF "ARGUMENT"

In western philosophy, an argument is traditionally defined as "a system composed of premisses and a conclusion" (Diogenes Laertius 7.45, citing a Stoic definition). The plural of "premisses" is a Stoic idiosyncrasy: few other philosophers have accepted the Stoic denial (Sextus Empiricus 2.443) that there are one-premissed arguments. But, even with an amendment to allow for one-premissed arguments, the definition is not very satisfactory, for two reasons: it requires a further explanation of what a premiss is and what a conclusion is, and it forecloses by stipulation

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

alternative conceptions of the components of an argument. It is therefore useful to develop a more informative and less question-begging conception, starting from everyday usage of the word "argument" and sharpening that usage so as to circumscribe a class of entities that is an appropriate subject for theoretical reflection.

In English, the word "argument" and the corresponding verb "argue" are used in two quite clearly distinguishable senses.

One sense is that in which we say such things as "John Searle argued that no computational system can have a semantics" or "the *Summa Theologica* of Thomas Aquinas contains five arguments for the existence of God". In this sense, arguing requires only one arguer (who in cases of collaboration in the production of an argument can be a group of people). The arguer expresses a point of view on a question, and offers as support for this position one or more reasons. The expression of the point of view and the provision of one or more reasons in its support constitute a complex of speech acts. The arguer addresses these speech acts to one or more readers, listeners or observers, who need not reply. Arguing in this sense is typically (though not always) emotionally neutral, and typically not accompanied by hostility.

The other sense is that in which we say such things as "they were arguing with one another" or "they had a bitter argument" or "she argued with him". In this sense, arguing requires at least two arguers; if one argues with oneself in this sense, then one sequentially takes two different roles. The arguers express to each other divergent opinions on some question. Each attempts to get the other(s) to accept their point of view, not necessarily by offering reasons in support of it. Emotional intensity and even hostility often accompany such disputes, though not always.

For ease of reference, I shall call these the reason-giving and the disputational senses of "argument" and "argue". Informal logic studies arguments in the reason-giving sense. It is worth noting that English is apparently unique in using the same word for these two senses. In classical Greek, for example, the reason-giving sense is expressed by the word *logos* (e.g. in Plato's *Phaedo*, at 90b–91c) in one of its many senses, whereas the disputational sense is expressed by the word *amphisbêtêsis* or *antilogia*, "dispute" or "controversy". In Latin, the reason-giving sense is expressed by the word *argumentum*, "proof" or "evidence", the disputational sense by the word *disputatio*, "debate" or "dispute". In French, as Plantin [2003, 174] points out in detail, the reason-giving sense is expressed by the verb *argumenter* ("to argue [that]") and its cognates, the disputational sense by the verb *discuter* ("to discuss", in an aggressive way). In Spanish [Claudio Duran, personal communication], the reason-giving sense is expressed by the word *argument*, and the disputational sense by the words *discusión* (discussion) or *controversia* (controversy) or *disputa* (dispute). In Russian, the reason-giving sense is expressed by the word *dovod* (supporting reason), the disputational sense by the word *spor* or *ssora.* In German, the reason-giving sense is expressed by the word *Argument*, the disputational sense by the word *Disput.* The reader is invited to check how other languages handle the distinction.

Reason-giving and disputation can of course coincide in a particular case. But not all reason-giving occurs in disputes; in fact, most of it occurs outside the context of disputes, or at least outside explicit ones. And not all disputes involve reason-giving; typically they do, but typically as well they involve other components that are intrinsic to the dispute. Productive disputation requires reason-giving. But if reason-giving and disputing were shown in an Euler diagram, the circles would overlap.

ARGUMENT AS DISCOURSE SUPPORTING A POINT OF VIEW BY OFFERING ONE OR MORE REASONS

An argument in the sense studied by informal logic can thus be conceived initially as a type of discourse in which the author expresses a point of view and offers one or more reasons in support of that point of view. We can make this conception more precise by considering in the first place what it is to offer a supporting reason for a point of view, and how one can do so. A typical means of doing so is to utter a sentence of a language, or something equivalent to a sentence. Searle [1979] has proposed a taxonomy of the illocutionary acts that people perform in uttering sentences, a classification based on differences in the point of the act.

Assertives have as their point to commit their utterers to the truth of an expressed proposition p; they include not only stating that p but also hypothesizing, suggesting, boasting, and deducing that p [Searle, 1979, 13]. *Directives* have as their point to get the addressee to do something; they include requesting, advising and asking [13–14]. *Commissives* have as their point to commit the utterer to do something; they include promising and contracting [8, 14]. *Expressives* have as their point to express a psychological attitude of the utterer to the state of affairs specified by the proposition expressed; they include apologizing, congratulating and thanking [15–16]. *Declaratives* have as their point to make something the case by the very utterance of the sentence; they including declaring war, christening, delivering judicial verdicts and stipulating how one is going to use a certain expression [16–20].

As Van Eemeren and Grootendorst point out [1984, 43], only an assertive (or something reconstructible as an assertive, such as a rhetorical question) can count as the offering of a supporting reason for a point of view. We can test this claim by considering examples of other sorts of illocutionary acts followed by "so" followed by an arbitrary sentence. If we consider a directive followed by "so" followed by a directive, such as:

* *What time is it? So you must go home.*

we find it difficult to make sense of such discourse, except by supposing that the speaker assumes that a correct answer to the question implies that it is past time for the addressee to go home. It is easier to make sense of a commissive followed by "so" followed by an assertive, such as:

> ? *I promise to pick up some milk on the way home. So you don't need to get it.*

But we can only make sense of such discourse because to make a commitment to do something in the future is implicitly to predict that one will do it. The argument would be more straightforward if one made the prediction explicitly by means of an assertive:

> *I will pick up some milk on the way home. So you don't need to get it.*

We strain to make sense of an expressive followed by "so" followed by an assertive, such as:

> * *Congratulations on your anniversary. So you are married.*

The act of congratulating a couple on their anniversary implies that they are married, but congratulating them is an awkward way of supporting the claim that they are married. It would be much more straightforward to say something like:

> *I congratulated them on their anniversary. So they are married.*

Similar problems arise in making sense of a declarative followed by "so" followed by an assertive:

> * *I hereby sentence you to two years less a day in prison. So the guards will now take you to prison.*

As with directives, we need to suppose an intermediate step in which the speaker expresses a commitment to the existence of the state of affairs brought about by the declarative:

> *You have just been sentenced to two years less a day in prison. So the guards will now take you to prison.*

Thus to offer a supporting reason by uttering a sentence, or something equivalent to a sentence, is to perform some sort of assertive, i.e. to commit the utterer to the truth of the expressed proposition. The word "truth" needs to be understood broadly as applying to normative and evaluative propositions as well as descriptive ones, for the following discourses make sense as arguments:

> *One must not cause unnecessary harm. So it is wrong to give someone distressing news that the person does not want to hear.*
>
> *All things considered, this car is the best model of the type we want that is in our price range. So let's buy it.*

It is possible, however, to offer a supporting reason without uttering a sentence. Drawings, figures, photographs, paintings, gestures, body language and other non-linguistic communicative devices can serve as premisses of an argument. Groarke

[1996] urges the recognition of visual arguments. Gilbert treats verbalized or verbalizable premiss-conclusion structures as one of four possible modes of argument, defined as "*any exchange of information centred on an avowed disagreement*" [Gilbert, 1997, 104]; italics in original). He calls this mode the "logical-critical" mode. Emotional arguments rely on the use and expression of emotion, and can be communicated without language [83]. Visceral arguments rely on physical activity, such as touching or body language [84]."Kisceral" arguments rely on the intuitive or the imaginative [86]. Although Gilbert uses the disputational rather than the reason-giving sense of argument, he gives convincing examples of reason-giving that is not verbalized. What is common to these various types of reason-giving is that their authors express to one or more addressees a commitment to the truth of a proposition.

Having specified a necessary condition for offering a supporting reason — namely, expressing to one or more addressees a commitment to the truth of a proposition — and indicated a variety of ways in which one may satisfy this necessary condition, we can gain further precision on the reason-giving sense of argument by considering what sorts of points of view can function as conclusions supported by the reason or reasons offered. Again, we can use the "so" test, this time focusing on what comes after the word "so" rather than what comes before it. Clearly we can express the conclusion of an argument by means of an assertive:

> *The whale suckles its young. So it is a mammal.*

But, as Pinto [2001a] points out, the arguer's endorsement of a point of view may be a directive, such as a request for information (*you were there, so what was it like?*) or a recommendation to do something (*there is a forecast of thundershowers, so let's cancel the picnic* or *I'm feeling cold, so please close the door*). It can also be a commissive (*I know how difficult it will be for you to get the milk, so I promise you that I will pick it up on the way home*), an expressive (*my conduct was inexcusable, so I apologize most sincerely*), or a declarative (*the evidence establishes beyond a reasonable doubt that you committed the crime of which you are accused, so I hereby find you guilty as charged*).

As with supporting reasons, conclusions can be expressed not only by uttering sentences of a language, or things equivalent to a sentence, but also by producing visual images, using "body language" including facial expressions, or performing physical actions like touching someone. Such non-verbalized conclusions are typically implicit in the communicative activity, and thus somewhat indeterminate.

ARGUMENTS AS INVITATIONS TO INFERENCE

What is crucial to an argument is the claim that the reasons collectively support the conclusion. The addressee of an argument is invited to accept the conclusion on the basis of the reasons offered. In Pinto's happy phrase, "Arguments are invitations to inference" [Pinto, 2001b, 37], where "inference" means "the mental act or event in which a person draws a conclusion from premisses, or arrives at a conclusion on the basis of a body of evidence" [32]. In inferring, a person adopts or reinforces an attitude towards the proposition embedded in the conclusion. These attitudes include a range of doxastic attitudes, from being convinced of it through being inclined to believe in it and suspecting it to considering it possible and having no idea about it [Pinto, 2001a, 12]. They also include such non-doxastic attitudes to propositions as fearing, desiring, intending and hoping [2001a, 16].

The condition that an argument is an invitation to an inference from the offered reasons to the conclusion applies even in suppositional reasoning and argument where the conclusion drawn shares the suppositional status of a premiss; the conclusion may in fact be an absurdity whose derivation will subsequently be used, in conjunction with an acknowledgement of its absurdity, to reject that premiss. It also applies, obviously, to arguing purely dialectically, from the assumptions of an interlocutor that one does not oneself share.

The claim that the offered reasons support the conclusion can be marked linguistically, by means of an illative expression governing the conclusion or a reason. Let us use the expression *premiss indicator* for an illative like "since" which (in its illative use) indicates that the immediately following assertive is offered in direct support of the speech act performed by uttering the main clause to which the "since" clause is subordinate, and the expression *conclusion indicator* for an illative like "therefore" which (in its illative use) indicates that the immediately preceding assertive is offered in direct support of the immediately following speech act. To introduce a reason by a premiss indicator is to perform a special type of assertive, which we might call premissing: to premiss a proposition is to put it forward as a (perhaps partial) basis for inferring a conclusion. Similarly, to introduce a conclusion by a conclusion indicator is to perform a special type of speech act (whether assertive, directive, commissive, expressive or declarative), which we might call concluding: to conclude a proposition is to put it forward for acceptance on the basis of one or more assertives offered as supporting reasons. Note that acceptance does not always mean adopting a doxastic attitude to the proposition; accepting an apology, for example, means believing that the apologizer bears some responsibility for the act for which the apology is offered and that this act was wrong and that the apologizer is sincerely sorry for this act, but it also means forgiving the apologizer, in the sense of not demanding further acts of contrition, reparation or penitence.

Arguments do not always include illatives, and even those that do include illatives typically attach them to only one component of the argument, a reason or the conclusion. But components of an argument not introduced by an illative are nevertheless premissed or concluded. Thus an argument is a claim-reason complex consisting of an act of concluding (which may be of any of the five main types in Searle's taxonomy of speech acts) and one or more acts of premissing (each of which is an assertive). These acts are correlative; the act of concluding is an act of concluding from the reasons, and each act of premissing is an act of offering support for the conclusion. To capture this relationship, it is appropriate to conceive of an argument as a sequence consisting of a set of reasons followed by a conclusion indicator followed by a conclusion, or equivalently as a conclusion followed by a premiss indicator followed by a set of one or more reasons. In such a sequence, the illative does the work of premissing each reason and concluding the conclusion; hence, we do not need to mention these acts in characterizing the reason and conclusion. Arguments with no explicit illative can be regarded as having one implicitly.

EXTENSIONS: POTENTIAL ARGUMENTS AND EQUIVALENCE CLASSES OF ARGUMENTS

So far I have been talking about actual arguments, actually advanced by people who speak, write or otherwise communicate them to one or more addressees. For the purposes of this chapter, I propose to extend the concept of argument further, in two respects. First, I propose to count as arguments discoursal claim-reason complexes that are merely entertained in thought, such as the pros and cons considered by people trying to come to a decision about what to do. The example fits ordinary usage, since we do talk about such people as considering the relevant arguments. But the general characterization goes beyond it, since it includes reasoning by oneself to a conclusion (*tomorrow is garbage day, so I had better put out the garbage*), whose content is not usually described as an argument. The reason for this extension is that the same considerations of acceptability of the supporting reasons and sufficiency of the support relation between reasons and conclusion apply to such solo reasoning as apply to other-directed arguments. Second, for the same reason, I propose to count as arguments merely potential discoursal claim-reason complexes never uttered or even mentally entertained by anyone.

In this extended sense, a simple argument is a sequence of three objects: a speech act c of any type concerning some proposition, an illative such as the word "since" (in its inferential sense), and a set P of one or more assertives. Expressed in this canonical form, the following would count as arguments:

> <express admiration for Picasso's *Guernica*, since, {assert that Picasso's *Guernica* brings home in a vivid way the horrible consequences for the innocent of aerial bombing in contemporary warfare}>
>
> <suspect that Goldbach's conjecture is correct, since, {assert that mathematicians have found no counter-example in 200 years of trying}>

These sequences may be purely possible ones, never articulated by anyone. The first could be expressed by saying, "What a wonderful painting is Picasso's *Guernica*. It brings home in a vivid way the horrible consequences for the innocent of aerial bombing in contemporary warfare." The second could be expressed by saying, "Goldbach's conjecture is probably correct, since mathematicians have found no counter-example in 200 years of trying."

General as it is, this definition is still less general than our ordinary usage of the word "argument". For the same argument can be expressed in different ways, and even in different languages. To accommodate this fact, we can extend the above definition to the equivalence class of all sequences with the same meaning as a given sequence. If $\langle c, \because, P\rangle$ is the given sequence, we may label its equivalence class $[\langle c, \because, P\rangle]$. It will include sequences with one or more constituents with the same meaning as the corresponding constituent of $\langle c, \because, P\rangle$, as well as corresponding sequences in the reverse order with a conclusion indicator in place of the premiss indicator, such as $\langle P, \therefore, c\rangle$. Then a simple argument may be defined as follows:

> *Simple argument* $=_{df}$ a class of those triples of the form $\langle c, \because, P\rangle$ or $\langle P, \therefore, c\rangle$ that are equivalent in meaning to one another, where c is a speech act of any type, $\because$ is a premiss indicator, $\therefore$ is a conclusion indicator, and P is a set of one or more assertives.

Following traditional terminology, we will refer to an assertive in such a set P as a *premiss* and to such a speech act c as a *conclusion.* Our usage is however non-traditional in this respect, that a premiss is neither a sentence nor a proposition nor a statement, but an assertive; and a conclusion is neither a sentence nor a proposition nor a statement, but a speech act of some type. Further, it should be noted that, in actual thinking, speaking, signing and writing, arguments often lack an inference indicator, the force of the illative being communicated by a combination of semantic and contextual factors.

The definition of a simple argument just proposed is vague, since it is indeterminate in some cases whether two linguistic expressions of an attitude have the same meaning. The vagueness is similar in degree to, and has the same source as, the vagueness in the concept of a proposition as the eternal object signified by the utterance of a sentence. A more precise conception of argument could be obtained by treating each linguistic variation in the formulation of an argument as a new argument. Someone who adopted this approach would confront vagueness at another place, in considering whether two distinct arguments have the same force.

COMPLEX DIRECT ARGUMENTS

So far we have been considering arguments in which one or more reasons are offered in direct support of a conclusion. A comprehensive conception of argument should allow for complex arguments, in which one or more of the reasons offered in direct support of a conclusion is in turn argued for. Such complex arguments can be analysed into component simple arguments. Anselm's ontological argument for the existence of God in the second chapter of his *Proslogium*, for example, is a tightly structured chain of inter-linked simple arguments. We can extend our definition to complex arguments, with the help of the concept of subordination, defined as follows:

> *Subordinate argument* $=_{df}$ an argument whose conclusion is a premiss of another argument.

Subordinate arguments are commonly referred to in the informal logic literature and textbooks as *subarguments.* We need also the concept of superordination, which is the converse relation:

> *Superordinate argument* $=_{df}$ an argument with a premiss that is the conclusion of another argument.

We can now define a complex argument as a set of two or more simple arguments, with a hierarchy of subordination between them:

> *Complex argument* $=_{df}$ a set of two or more simple arguments, each of which is either superordinate to or subordinate to at least one other argument in the set, and one of which (the *main argument*) is not subordinate to any other argument in the set.

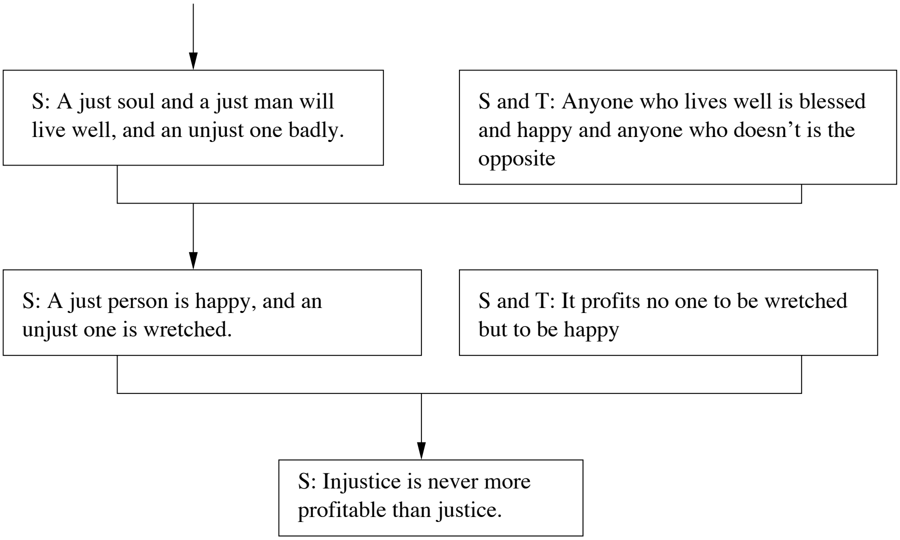
We can illustrate this definition with the concluding part of the first book of Plato's *Republic*, where Socrates argues, with the agreement of Thrasymachus at each step, that injustice is never more profitable than justice:

> . . . a just soul and a just man will live well, and an unjust one badly.
> Apparently so, according to your argument.
> And surely anyone who lives well is blessed and happy, and anyone who doesn't is the opposite.
> Of course.
> Therefore, a just person is happy, and an unjust one wretched.
> So be it.
> It profits no one to be wretched but to be happy.
> Of course.
> And so, Thrasymachus, injustice is never more profitable than justice.
> Let that be your banquet, Socrates, at the feast of Bendis.
>
> (Plato, *Republic* I. 353e-354a, Grube-Reeve translation)

This excerpt is the concluding part of a single extremely complex argument, in which attitudes are expressed to about 50 different propositions. The excerpt can be represented in the standard notation for sets and sequences as follows:

> {<{assert that a just soul and a just man will live well and an unjust one badly, assert that anyone who lives well is blessed and happy and anyone who doesn't is the opposite}, ∴, assert that a just person is happy, and an unjust one wretched>, <{assert that a just person is happy and an unjust one wretched, assert that it profits no one to be wretched but to be happy}, ∴, assert that injustice is never more profitable than justice>}

In a standard box-arrow system of diagrammatic representation, the argument is as follows:


S: A just soul and a just man will live well, and an unjust one badly.
S and T: Anyone who lives well is blessed and happy and anyone who doesn't is the opposite
S: A just person is happy, and an unjust one is wretched.
S and T: It profits no one to be wretched but to be happy
S: Injustice is never more profitable than justice.

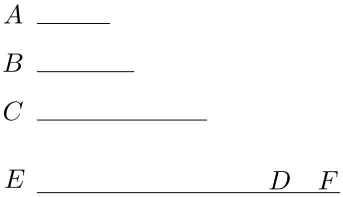
Thus this text is a complex of two simple arguments, each with two premisses, with the conclusion of the subordinate argument identical to one premiss of the main argument. Both Socrates and Thrasymachus independently affirm each of the ultimate premisses. It is noteworthy that Thrasymachus gives only grudging acknowledgement of the conclusions that Socrates draws in each simple argument, including the conclusion of a previous argument with which the quoted excerpt begins. It is also noteworthy that the argument is reported, not advanced in his own name by its author (Plato). This fact adds an additional level of complexity to the text.

SUPPOSITIONAL ARGUMENTS

The preceding definition covers direct arguments to a conclusion. It needs to be expanded to cover as well *reductio ad absurdum* arguments, suppositional arguments with a conditional conclusion, and other arguments in which a conclusion is drawn on the basis of an argument. The easiest way to accommodate such arguments is to expand the concept of a premiss so that it includes not only assertives whose content is a proposition but also arguments (which are complexes of illocutionary acts). Consider for example Euclid's proof that there are more than any given [finite] number of primes:

> The prime numbers are more numerous than every given number of prime numbers.

> Let the given prime numbers be A, B, C. I say that there are more prime numbers than A, B, C.
> For let the least number measured [i.e. divisible — DH] by A, B, C be taken and let it be DE, and let a unit DF be added to DE. Then EF either is prime or not.



> First, let it be prime. Therefore A, B, C, EF are prime numbers discovered to be more numerous than A, B, C.
>
> Next, let EF not be prime. Therefore it is measured by some prime number. (VII. 31) Let it be measured by the prime number G. I say that G is the same as none of A, B, C. For if it is possible, let it be so. A, B, C measure [i.e. are factors of — DH] DE, and therefore G will measure DE. And it measures EF. And being a number G will measure the unit DF, which is absurd. Therefore G is not the same as one of A, B, C. And therefore it is supposed that G is not the same as one of A, B, C; and it is supposed prime. Therefore A, B, C, G are prime numbers discovered to be more numerous than A, B, C; which it was necessary to prove. (Euclid, *Elements* IX.20; translation by the present author)

Euclid's proof is a simple argument with several layers of embedding. It can be represented in the standard notation for sets and sequences as follows:

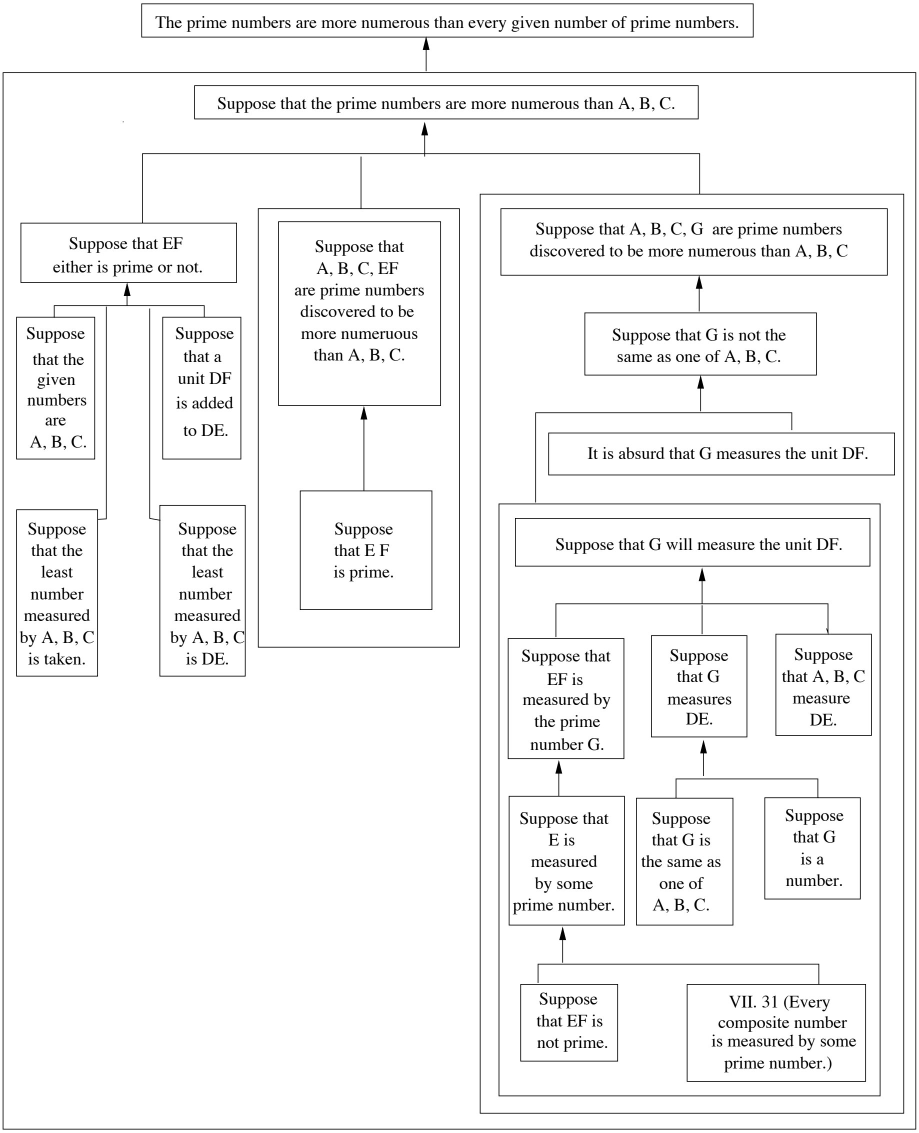
> <assert that the prime numbers are more numerous than every given number of prime numbers, $\because$, {<{suppose that the given prime numbers are A, B, C; suppose that the least number measured by A, B, C is taken; suppose that the least number measured by A, B, C is DE; suppose that a unit DF is added to DE}, $\therefore$ suppose that EF either is prime or not>, <{suppose that EF either is prime or not, <{suppose that EF is prime}, $\therefore$, suppose that A, B, C, EF are prime numbers discovered to be more numerous than A, B, C >, < {{ <{suppose that EF is not prime, assert VII.31}, $\therefore$, suppose that EF is measured by some prime number>, <{suppose that EF is measured by some prime number}, $\therefore$, suppose that EF is measured by the prime number G >, <{suppose that G is the same as one of A, B, C, suppose that A, B, C measure DE}, $\therefore$, suppose that G will measure DE >, <{suppose that EF is measured by the prime number G, suppose that G will measure DE, suppose that G is a number}, $\therefore$, suppose that G will measure the unit DF >}, assert that it is absurd that G measures the unit DF}, $\therefore$,

> suppose that G is not the same as one of $A, B, C >$, <suppose that G is not the same as one of $A, B, C\}$, ∴, suppose that A, B, C, G are prime numbers discovered to be more numerous than $A, B, C >\}\}$, ∴, suppose that the prime numbers are more numerous than $A, B, C >\}\}>$

The outermost argument supports a universal generalization with an argument that an arbitarily chosen instance has the property of interest. This embedded argument is an argument by cases, whose premisses are a supposition about what the two cases are (supported by a sub-argument) and two embedded arguments reaching the desired conclusion for each case. The argument for the first case is very simple. The second is complex, with a main premiss supported by a *reductio ad absurdum* argument which is the heart of the proof. We can diagram the whole argument as in the diagram on the next page.

Thus allowing arguments to be premisses permits one to represent in a uniform way generalizations from a result proved for an arbitrarily chosen instance, proofs by cases, and *reductio ad absurdum* arguments. Similarly one can represent proofs by conditional reasoning, as in Saint Anselm's ontological argument for the existence of God:

> **Truly there is a God, although the fool has said in his heart, There is no God.**
>
> AND so, Lord, do you, who do give understanding to faith, give me, so far as you knowest it to be profitable, to understand that you are as we believe; and that you are that which we believe. And indeed, we believe that you are a being than which nothing greater can be conceived. Or is there no such nature, since the fool has said in his heart, there is no God? [Psalms xiv, 1]. But, at any rate, this very fool, when he hears of this being of which I speak — a being than which nothing greater can be conceived — understands what be hears, and what he understands is in his understanding; although he does not understand it to exist.... Hence, even the fool is convinced that something exists in the understanding, at least, than which nothing greater can be conceived. For, when he hears of this, he understands it. And whatever is understood, exists in the understanding.
>
> And assuredly that, than which nothing greater can be conceived, cannot exist in the understanding alone. For, suppose it exists in the understanding alone: then it can be conceived to exist in reality; which is greater. Therefore, if that, than which nothing greater can be conceived, exists in the understanding alone, the very being, than which nothing greater can be conceived, is one, than which a greater can be conceived. But obviously this is impossible.
>
> Hence, there is no doubt that there exists a being, than which nothing greater can be conceived, and it exists both in the understanding and in reality. ([Anselm, *Proslogium* 2], Deane's translation; paragraphing altered to facilitate understanding)


The prime numbers are more numerous than every given number of prime numbers.
Suppose that the prime numbers are more numerous than A, B, C.
Suppose that EF either is prime or not.
Suppose that the given numbers are A, B, C.
Suppose that a unit DF is added to DE.
Suppose that the least number measured by A, B, C is taken.
Suppose that the least number measured by A, B, C is DE.
Suppose that A, B, C, EF are prime numbers discovered to be more numeruous than A, B, C.
Suppose that E F is prime.
Suppose that A, B, C, G are prime numbers discovered to be more numerous than A, B, C
Suppose that G is not the same as one of A, B, C.
It is absurd that G measures the unit DF.
Suppose that G will measure the unit DF.
Suppose that EF is measured by the prime number G.
Suppose that G measures DE.
Suppose that A, B, C measure DE.
Suppose that E is measured by some prime number.
Suppose that G is the same as one of A, B, C.
Suppose that G is a number.
Suppose that EF is not prime.
VII. 31 (Every composite number is measured by some prime number.)

Anselm's argument can be represented in the notation of sets and sequences as follows: {<{assert that God is a being than which nothing greater can be conceived, assert that the fool has said in his heart there is no God, assert that when he hears of this he understands it, assert that whatever is understood exists in the understanding}, ∴, assert that something exists in the understanding, at least, than which nothing greater can be conceived>, {<{<{suppose that that than which nothing greater can be conceived exists in the understanding alone}, ∴, suppose that that than which nothing greater can be conceived can be conceived to exist in reality as well as in the understanding>, assert that it is greater to exist in reality than to exist in the understanding alone}, ∴, assert that if that than which nothing greater can be conceived exists in the understanding alone the very being than which nothing greater can be conceived is one than which a greater can be conceived>, <{assert that if that than which nothing greater can be conceived exists in the understanding alone the very being than which nothing greater can be conceived is one than which a greater can be conceived, assert that it is impossible that the very being than which nothing greater can be conceived is one than which a greater can be conceived}, ∴, assert that that than which nothing greater can be conceived cannot exist in the understanding alone>}, <{assert that something exists in the understanding, at least, than which nothing greater can be conceived, assert that that than which nothing greater can be conceived cannot exist in the understanding alone}, ∴, assert that there exists a being than which nothing greater can be conceived and it exists both in the understanding and in reality>}.

Diagrammatically, Anselm's argument looks as in the diagram on the next page.

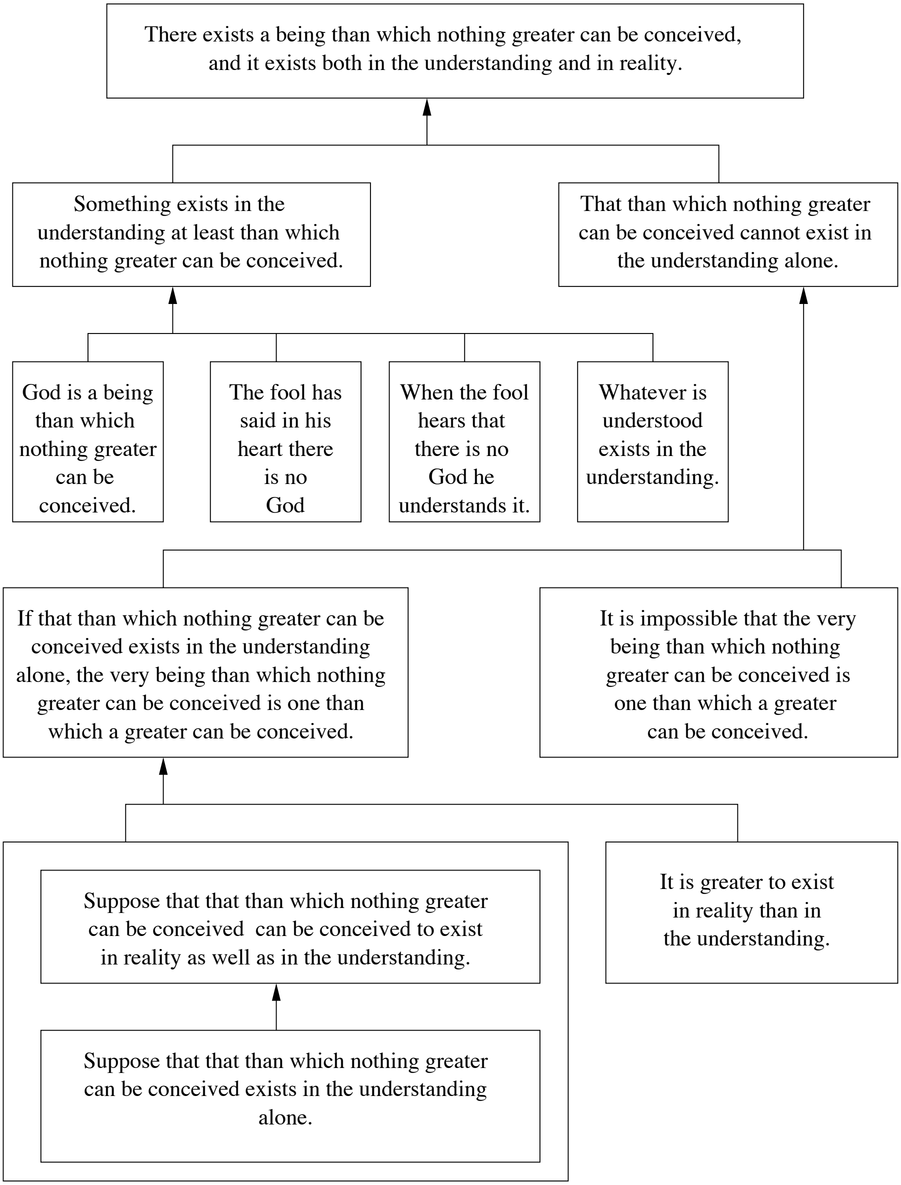
The main argument has two premisses, each an assertion. The first main premiss is supported by a simple argument with four premisses, each an assertion. The second main premiss is supported immediately by two premisses, one an assertion of a conditional and the other an assertion of the impossibility of its consequent. The assertion of the conditional is in turn supported by two premisses, one of them an argument from the supposition of the antecedent of the conditional to a suppositional conclusion and the other an assertion.

The approach of allowing arguments as premisses also permits one to represent proofs by mathematical induction in the same uniform way.

FIRST SUMMARY

In summary, one can integrate the foregoing considerations into a single recursive definition of an argument, as follows:

1. Any set of the form $\{\langle c, \because, P\rangle\}$ or $\{\langle P, \therefore, c\rangle\}$ is an argument, where the conclusion c is a speech act of any type, $\because$ is a premiss indicator, $\therefore$ is a conclusion indicator, and the set P of premisses is a set of one or more assertives.


There exists a being than which nothing greater can be conceived, and it exists both in the understanding and in reality.
Something exists in the understanding at least than which nothing greater can be conceived.
That than which nothing greater can be conceived cannot exist in the understanding alone.
God is a being than which nothing greater can be conceived.
The fool has said in his heart there is no God
When the fool hears that there is no God he understands it.
Whatever is understood exists in the understanding.
If that than which nothing greater can be conceived exists in the understanding alone, the very being than which nothing greater can be conceived is one than which a greater can be conceived.
It is impossible that the very being than which nothing greater can be conceived is one than which a greater can be conceived.
Suppose that that than which nothing greater can be conceived can be conceived to exist in reality as well as in the understanding.
Suppose that that than which nothing greater can be conceived exists in the understanding alone.
It is greater to exist in reality than in the understanding.

2. Any set equivalent in meaning to a set of the form described in clause 1 is an argument.

3. If a conclusion in an argument A is a premiss in an argument B, then $A \cup B$ is an argument.

4. If $\{\langle P, \therefore, c\rangle\}$ is an argument, and A is an argument, then so are $\{\langle A \cup P, \therefore, c\rangle\}$ and $\{\langle A, \therefore, c\rangle\}$. Similarly for $\{\langle c, \because, P\rangle\}$.

5. Nothing is an argument unless it can be constructed in a finite number of steps using the above rules.

COMPETING CONCEPTIONS OF ARGUMENT

To appreciate the force of the preceding definition, one can compare it to recent thematic discussions of the concept of argument.

Charles Hamblin in his classic chapter "The concept of argument" [1970, 224–252] develops an account of what an argument is from the generally agreed conception of an argument as "whatever it is that is typically expressed by the form of words 'P, therefore Q', 'P, and so Q', 'P, hence Q'; or, perhaps, 'Q, since P', 'Q, because P"', and cites Whately's *Elements of Logic* [1848] for the standard terminology of "premisses" for whatever is expressed by P in such forms and "conclusion" for whatever is expressed by Q [228]. He approaches indirectly the task of saying what is typically expressed by such forms of words, through the question of what constitutes a good argument. This question he addresses dialectically, first enunciating a set of "alethic" criteria based on the concept of truth, then overcoming their inadequacy by shifting to a set of epistemic criteria based on the concept of knowledge, and then in turn overcoming the inadequacy of the epistemic criteria by shifting to the following set of dialectical criteria based on the concept of acceptance (by a single person or by a group of persons):

> (D1) *The premisses must be accepted.*
> (D2,3) *The passage from premisses to conclusion must be of an accepted kind.*
> (D4) *Unstated premisses must be of a kind that are accepted as omissible.*
> (D5) *The conclusion must be such that, in the absence of the argument, it would not be accepted.* [Hamblin, 1970, 245], italics in original)

Hamblin does not say in so many words what account of argument these criteria imply. If we assume that the criteria for a good argument, like the criteria for a good pruning knife or for a good eye, are those that are jointly sufficient and individually necessary for it to do its work well, then Hamblin's criteria imply that an argument has as its function to secure acceptance of its conclusion by its addressee(s) on the basis of acceptance of its premisses. Thus we may attribute to Hamblin a definition of an argument as an attempt to get one or more people to accept an expressed conclusion Q by adducing in its support some premisses P. This definition is more restrictive than the conception advanced by the present author in this chapter, in at least three respects. First, it covers only simple arguments, not complex ones. Second, it does not explicitly allow an argument to be a premiss. Third, and most importantly, it restricts arguments to a single function: considerations for and against a certain decision that one rehearses to oneself do not count as arguments, nor is there any allowance for arguments designed to prove or justify their conclusion, where acceptance is not enough and perhaps not even necessary. Hamblin's definition differs from the preceding conception in at least two additional respects. Rather curiously, he does not treat a linguistic communication of the form "P, therefore Q" as an argument; rather, an argument for him is what such a form of words expresses, i.e. some sort of abstract object, whose components and ontological status he does not specify. In addition, he makes no allowance for modal qualification of premisses and conclusions, e.g. in suppositional arguments.

Douglas Walton [1990, 411] proposes to define argument as

> a social and verbal means of trying to resolve, or at least contend with, a conflict or difference that has arisen between two (or more) parties. An argument necessarily involves a claim that is advanced by at least one of the parties.

This definition fits the disputational sense of "argument" much better than the reason-giving sense. It includes among arguments verbal exchanges of conflicting claims (A: *Yes, you did.* B: *No, I didn't.*) in which the parties give no supporting reasons. It excludes written texts whose authors provide reasons to support a claim that nobody has ever questioned or rejected, for example, mathematical proofs of theorems which nobody has previously thought about. It also excludes discourses in which a speaker makes a case for a claim that the listeners already accept, as a way of reinforcing the adherence of the audience to the claim (as in preaching to the converted). There may be a point to focusing theoretically on verbal expressions of conflict, but it would be less confusing to use some other word than "argument" for them. Perhaps "disputation" would be a better word.

Robert Pinto [2001b], as already noted, distinguishes inference, "the mental act or event in which a person draws a conclusion from premisses, or arrives at a conclusion on the basis of a body of evidence" [32], from argument, initially characterized as "a set of statements or propositions that one person offers to another in the attempt to induce that other person to accept some conclusion" [32]. The success of such an attempt, he points out, is not just a matter of its causing its addressee(s) to accept its conclusion; for example, if someone concedes the conclusion merely out of exhaustion from an argument's overwhelming length and complexity, then the argument has not persuaded this addressee to accept its conclusion. What is crucial is that the addressee make the mental inference that corresponds to the move in the argument from premisses to conclusion. Thus "arguments are invitations to inference" [Pinto, 2001b, 37]. This conception of argument has been central to the development of the definition proposed in this chapter. But Pinto's conception differs from the present definition in that it treats arguments as products with a particular purpose (to get the addressee to accept the conclusion) rather than as abstract objects that might be used for various purposes.

Ralph Johnson proposes the following definition:

> An argument is a type of discourse or text — the distillate of the practice of argumentation — in which the arguer seeks to persuade the Other(s) of the truth of a thesis by producing the reasons that support it. In addition to this illative core, an argument possesses a dialectical tier in which the arguer discharges his dialectical obligations. [Johnson, 2000, 168]

The practice of argumentation to which Johnson refers in this definition he defines as "the sociocultural activity of constructing, presenting, interpreting, criticizing, and revising arguments." [Johnson, 2000, 12, 154] The present author has argued, and Johnson has apparently accepted [Johnson, 2002, 313], that his intention is best captured by adding to this definition of the practice the qualifying phrase "for the purpose of reaching a rationally shared position on some issue" [Hitchcock, 2002, 291]. Further, to avoid the circularity of defining arguments in terms of argumentation and argumentation in terms of arguments, Johnson must recognize explicitly in his definition of argument (as he does implicitly elsewhere in his book) that arguments occur in contexts other than that of argumentation. The following revision of Johnson's definition of argument, intended to meet this and other difficulties, is one that Johnson apparently accepts [Johnson, 2002, 313]:

> An argument is a spoken discourse or written text whose author (the arguer) seeks to persuade an intended audience or readership (the Other or the Others) to accept a thesis by producing reasons in support of it. In addition to this illative core, an argument possesses a dialectical tier in which the arguer discharges his dialectical obligations. [Hitchcock, 2002, 289]

The dialectical obligations to which Johnson refers are obligations to address objections to and criticisms of the illative core and to consider alternative positions. In response to criticism of his controversial requirement that arguments must have a dialectical tier, Johnson has clarified his position: not all arguments have a dialectical tier, but "the *paradigm* case of an argument — the one that we should base our theories on and make policies over — is that in which there is both illative core and dialectical tier" [Johnson, 2002, 316]. The definition of argument proposed in the present chapter clearly differs from Johnson's definition in treating arguers' discharge of their dialectical obligations as extrinsic to their actual argument. The exclusion of such dialectical material from arguments proper conforms to our ordinary usage of the term "argument", and is compatible with Johnson's insistence that arguers have a responsibility to discharge their dialectical obligations. It might be better to use the word “case” for the whole complex consisting of an argument and its dialectical penumbra.

Johnson’s definition differs from the definition proposed in this chapter in at least two additional respects. It restricts arguments to actual discourses or texts rather than considering them as abstract objects that may be unexpressed. And it requires that their authors have as their purpose to persuade an intended recipient to accept a thesis on the basis of the reasons supplied, whereas the definition proposed in the present chapter leaves undetermined the purpose for which someone might express an argument. In defence of the latter position, one might point out that it is possible to include an argument as part of a joke, or that scholars and scientists who make a case for some position in academic writing may be more concerned to get on the record a solid justification of the position than to actually persuade any particular person or persons to accept it. In general, then, there is a variety of purposes for which people express arguments.

J. Anthony Blair [2004] construes arguments as reasons for something: for beliefs or for believing, for attitudes or for emotions, or for decisions about what to do. A set of propositions is a reason for something if and only if they actually support it. Blair justifies his abstract conception by pointing out that arguments as we ordinarily understand them have many uses — not just persuasion, but also quasi-persuasion, inquiry, deliberation, justification, collaboration, rationale-giving, edification, instruction and evaluation. Hence it distorts our ordinary understanding of arguments to build into the definition of argument some particular purpose such as persuasion or the resolution of a conflict of opinion.

Blair’s general approach of treating an argument as an abstract object that can be used for various purposes corresponds to the approach of the present chapter, as does his catholicity about the types of objects that can serve as conclusions of an argument. It differs from the approach of the present chapter in that it requires that the premisses of an argument actually support its conclusion, and indeed that none of the premisses is superfluous. He recognizes that people sometimes offer as a reason some consideration that does not in fact support the conclusion drawn from it. On his view, such people take themselves to be advancing an argument, but in fact they are not doing so. The term “argument” thus acquires a normative force, something like the force that the term “art” or “music” has in some people’s usage. Blair’s restriction of arguments to structures in which the premisses actually support their conclusion clearly departs from our actual use of the term “argument”. Further, it is theoretically awkward, since informal logic deals with the identification, analysis, evaluation and criticism of inferentially bad arguments as well as inferentially good ones. What name then will Blair use for its subject-matter? The restriction of arguments to inferentially good ones appears to be motivated by Blair’s treatment of an argument as a set, one of whose members is the conclusion and the rest of which are the premisses. Without a requirement that the premisses actually support the conclusion, every set of propositions would count as an argument — a classification clearly at odds with our ordinary usage.

The present chapter avoids such an overly broad conception by treating a (simple) argument as a sequence, with an illative as well as premisses and conclusion. It also differs from Blair's conception in allowing complex arguments and in allowing an argument to be a premiss.

EXCLUSIONS FROM THE CLASS OF ARGUMENTS

Further clarification of the force of the definition of argument proposed in the present chapter should emerge from a consideration of the sorts of objects that the definition excludes as not being arguments.

First, not all persuasive communication counts as an argument in the sense defined. Students of persuasion commonly accept the distinction already pointed by Aristotle (*Rhetoric* I.2.1356a1-20) of a variety of means of persuasion: character, emotion, argument (in Greek *êthos*, *pathos*, *logos*). Presentation of oneself as having a certain character may enhance the credibility of what one says, but it is not an argument in the sense defined in the present chapter, since it lacks a premiss-conclusion structure. For the same reason, stirring up the emotions of one's audience is not in itself an argument, even though it may be more effective than argument at moving them, and even though it can be combined with argument. Talk of "ethotic argument" and "pathetic argument" blurs an important distinction among different means of persuasion. Similarly, not all advertising contains arguments. Some advertising, for example legal advertising, is purely informative. But even persuasive advertising often works by presenting the product or service in an appealing manner, by creating associations through visual imagery, or by providing detailed information about it. Only advertising with an explicit premiss-conclusion structure counts as an argument according to the definition of the present chapter.

Second, insinuation is not argument. Although someone who insinuates something invites the hearer or reader to draw a conclusion from their words, the words themselves do not draw that conclusion. One can of course identify and discuss the argument that the insinuator invites us to construct for ourselves. Typically, however, this argument is rather indeterminate, precisely because insinuation merely suggests.

Third, some visual, emotional, visceral and kisceral communication is argumentative. But some is not, even if it has a persuasive function. Images can supplement written words as part of an argument. The drawings of lines that accompany Euclid's proof that there are more primes than any given number of primes enhance one's understanding of the suppositions he makes and of the construction that is at the heart of the proof; they can be interpreted as a visual repetition of what is written in words. A diagram of an experimental apparatus serves a similar function of supplementing a parallel verbal description in the methods section of a scientific paper. Visual images can also function as ineliminable components of an argument. A poster with a giant photograph of a starving emaciated child and the words "make poverty history" can reasonably be construed as an argument in the sense defined in this chapter; its premiss is the situation exhibited in the photograph, whose information content has components that no purely verbal description can supply, and its conclusion is the directive of the written text. But visual images that merely suggest a conclusion to be drawn by the viewer are not arguments in the sense defined in this chapter, because they do not contain a conclusion. Similarly, emotional expressions can have a persuasive effect without being part of an argument; for example, a conciliatory tone of voice may help persuade someone to calm down, but it is not part of an argument unless the speaker in so many words urges the addressee to calm down and states some reason for her or him to do so. Likewise with visceral communication; a threatening gesture may persuade someone to accede to some implied request, but it becomes part of an argument only when the gesturer says, “Your money or your life”, or words to that effect. As for Gilbert’s “kisceral” mode, it is debatable whether the intuitions or hunches to which it refers are communicative devices, as opposed to a type of evidence analogous to direct observation. Of course, persons may cite their intuitive feeling as a reason for a certain belief or decision, but in that case the premiss is the assertion that they have this feeling, rather than the feeling itself.

SECOND SUMMARY

To sum up, this chapter has proposed a definition of an argument as a set of one or more interlinked premiss-illative-conclusion sequences. Such sequences can be interlinked either through chaining together, when the conclusion of one sequence is a premiss of another, or through embedding, when one sequence is a premiss of another. A premiss is an assertive, conceived as not necessarily asserted by anyone, and a conclusion is a speech act of any type, conceived as not necessarily performed by anyone or urged upon any addressee. In other words, arguments are abstract structures. When expressed, whether in language or in images or in physical behaviour, an argument invites its addressees to accept each conclusion on the basis of the acceptance of the assertives in its immediately supporting reasons.

POSTSCRIPT

The following is a list of some questions about arguments investigated within informal logic, with some references to the relevant literature:

1. *On argument identification*

How can one determine whether there is an argument in a spoken discourse, written text or other human communication?

In particular, what is the difference between an argument and a causal explanation? How can one tell in particular cases whether an indicator word like *because* or *hence* is being used inferentially or causally or both?

2. On argument analysis

What are the components of an argument? The standard view, incorporated in the definition of argument in this chapter, is that a simple argument has two types of components, the premisses and a conclusion, possibly linked by an illative. An alternative to this view, widely adopted in the field of speech communication, is the model of Stephen Toulmin for the "layout of arguments" [Toulmin, 1958], according to which a simple argument has six components: claim, data (later grounds), warrant, backing, modal qualifier, rebuttal. Another proposal, advanced by Rolf George [George, 1983] on the basis of Bolzano's conception of consequence in his 1837 *Wissenschaftslehre*, is that a fully specified simple argument has three components: premisses, conclusion, and variands; the variands are those parts of the premisses and conclusion that are subject to variation in determining whether the conclusion is a consequence of the premisses. Still another proposal, due to Ralph Johnson [2000], is that an argument has not only the premisses and conclusion of its "illative core" but also a dialectical tier in which the arguer responds to objections to and criticisms of the illative core and addresses alternative positions.

In what ways can two or more premisses offer direct support to a single conclusion? What test should be used to determine how the premisses of a multi-premiss simple argument support the conclusion? Beardsley introduced the concept of a *convergent argument* as one where "several independent reasons support the same conclusion" [Beardsley, 1950, 19]. For Beardsley, however, a reason could consist of several premisses working together. To mark an argument with such a multi-premiss reason, Thomas introduced the concept of a *linked argument*, defined as one that "involves several reasons [i.e. premisses — DH], each of which is helped by the others to support the conclusion" [Thomas, 1973]. Although the distinction seems intuitively clear, different authors have proposed different tests for determining whether a multi-premiss simple argument is linked or convergent, and these tests give different results from one another and in some cases from our intuitive judgements about particular arguments. Walton [1996a, 109–150] distinguishes five types of tests and skilfully displays the different classifications that they produce for a number of cases of argument and the difficulties for each type of test. For simplicity, he defines these tests for a two-premiss simple argument, but the definitions can easily be extrapolated to simple arguments with more than two premisses. The (necessary and sufficient) conditions for a two-premiss argument to be linked are as follows:

Falsity/no support: Each premiss by itself gives no support to the conclusion if the other premiss is false.

Suspension/insufficient proof: Each premiss by itself gives insufficient support to prove the conclusion if the other premiss is suspended, i.e. taken as not proved or not known to be true.

Falsity/insufficient proof: Each premiss by itself gives insufficient support to prove the conclusion if the other premiss is false.

Suspension/no support: Each premiss by itself gives no support to the conclusion if the other premiss is suspended, i.e. taken as not proved or not known to be true.

Degrees of support: The premisses together make the overall strength of the argument much greater than they would considered separately.

A sixth test not mentioned by Walton is due to Vorobej [1994]:

Type reduction upon elimination (TRUE): The argument is of a different type, with weaker support for the conclusion, if a premiss is eliminated.

A particularly difficult type of argument to classify as linked or convergent is what Walton [1996a, 130–134] calls an "evidence-accumulating" argument, where each premiss by itself gives some support to the conclusion but the combination of premisses gives more support. An example is the accumulation of symptoms and signs supporting a physician's diagnosis of a particular patient. On some tests such arguments come out linked, on others convergent. Some authors (e.g. Snoeck Henkemans 1992) distinguish the *cumulative* support exhibited by such arguments as a third type distinct from either linked or convergent support. Vorobej [1995] notes that some two-premiss simple arguments come out linked if a given test is applied to one premiss and convergent if the test is applied to the other premiss; he calls these *hybrid arguments*.

What standard forms can be used to represent the structure of complex argumentation? One method, carried out in detail by Maurice Finocchiaro in his analysis of Galileo's *Dialogue Concerning the Two Chief World Systems* [Finocchiaro, 1980] and found with variations in many textbooks, is to use a numbering system that indicates the support relationships claimed in the text — for example, C for the main conclusion; 1, 2 and so on for premisses offered in direct support of C; 1.1, 1.2 and so on for premisses offered in direct support of 1; and so on. Such numbered components can be indented to exemplify visually the support relationships, as in the following "standardization" of the argument quoted earlier in this chapter from the end of Book I of Plato's *Republic*:

        1.1 A just soul and a just man will live well, and an unjust one badly.
        1.2 Anyone who lives well is blessed and happy, and anyone who doesn't is the opposite.
    1. Therefore, a just person is happy, and an unjust one wretched.
    2. It profits no one to be wretched but to be happy.
C. Therefore, injustice is never more profitable than justice.

Such numbering systems can be extended to accommodate structures where an argument is a premiss. Another method is to use diagrams such as the box-arrow diagrams used in the present chapter or the diagrams in Toulmin's *The Uses of Argument* [1958]. Software is available for constructing such diagrams: for example, Araucaria [Reed and Rowe, 2005], Athena Standard [Rolf and Magnusson, 2002] and Reason!able [Van Gelder, 2004].

What principles should be followed in extracting arguments from human communications and putting them in a standard form or diagram? In what respects can an analyst alter the content of an argument component, and why? What components in the text, e.g. repeated components, can an analyst delete, and why? What components can an analyst add, and why? In particular, under what circumstances does a communicated argument have an unstated "gap-filling" [Ennis, 1982] premiss which the analyst can add? How is it to be determined what exactly is the unstated premiss in such a case?

Can the traditional division of arguments into deductive arguments and inductive arguments be defended, and if so on what basis? Are there arguments that fit into neither category, such as arguments by analogy, balance-of-considerations arguments (also known as pros and cons reasoning or conductive arguments [Wellman, 1971; Govier, 1987]), means-end reasoning, abductive reasoning (also known as inference to the best explanation), and the dozens of other types distinguished in the literature on argumentation schemes [Perelman and Olbrechts-Tyteca, 1958; 1969; Ehninger and Brockriede, 1963; Hastings, 1963; Van Eemeren and Grootendorst, 1992, 94–102; Kienpointner, 1992; Walton, 1996b; Grennan, 1997, 151–219]?

An influential tradition within informal logic construes arguments as advanced in dialogue, even when there is no actual intervention by an interlocutor. What types of dialogues are there? What is the function of each type, and what rules govern its participants? Hamblin [1970] proposed a discipline of "formal dialectic", within which he thought that the fallacies tradition could be made intellectually respectable. Van Eemeren and Grootendorst [1984; 1992; 2004] construe all arguments according to the normative model of what they call a "critical discussion". Walton and Krabbe [1995, 66] distinguish six main pure types of dialogue, each with its own goals: persuasion, negotiation, inquiry, deliberation, information-seeking, eristic. They propose detailed rules for two sub-types of persuasion dialogues: permissive persuasion dialogues and restrictive persuasion dialogues [Walton and Krabbe, 1995, 123–172].

3. On argument evaluation

What are the criteria for a good argument? Hamblin [1970, 224–252] usefully distinguishes alethic, epistemic and dialectical criteria. To these alternatives, one should add rhetorical criteria, focused on effective persuasion [Wenzel, 1980].

What objections can be raised to alethic criteria? Is truth of an argument's premisses even a necessary condition for it to be a good argument, let alone a sufficient one?

What objections can be raised to epistemic criteria? Is there a defensible set of epistemic criteria for a good argument? One epistemic approach, adopted by Goldman [1999], is to adopt a "veritistic" criterion for the goodness of arguments, according to which arguments are good insofar as they tend to lead to true conclusions; this approach of course applies only to arguments whose conclusions are assertions. Another epistemic approach, adopted among others by Feldman [1994] and by Siegel and Biro [1997], is to adopt a justificatory criterion, according to which an argument is good insofar as its premisses justify its conclusion.

What objections can be raised to dialectical criteria? Does mere acceptance by an interlocutor of an argument's starting-points and inferences make the argument a good one?

What objections can be raised to rhetorical criteria? Can one defend a basically rhetorical approach to the evaluation of arguments?

On the definition of argument proposed in the present chapter, an argument is an abstract structure that can be used for different purposes. If the evaluation of an artefact is relative to the purpose for which it is being used, then the evaluation of an argument will also be relative to its use in a given context. What are the different uses to which human beings put arguments, and what criteria for a good argument does each such use imply [Blair, 2004]?

On the argumentation schemes approach, each argumentation scheme has associated with it a set of "critical questions" that must be answered positively in order for an argument conforming to that scheme to be a good argument. How are these critical questions determined? What are the critical questions for each argumentation scheme? Does a satisfactory answer to the critical questions for a given argumentation scheme imply that the argument in question conclusively establishes its conclusion? Or is there still room for defeat of the argument by further information? For a given argumentation scheme, to what extent is there a burden on the author of an argument conforming to that scheme to show that there are positive answers to the critical questions for that scheme?

What types of defeaters are there? How is the status of an argument to be adjudicated as a sequence of various types of defeaters, defeaters of defeaters, and so on, is noticed?

To what extent is the author of an argument obliged to consider in a "dialectical tier" objections, criticisms and alternative positions?

What is a fallacy? As Hamblin [1970] correctly reports, the logical tradition assumes that it is a type of argument that merely seems valid. Hamblin himself proposed that the study of fallacies be made part of a new discipline that he called "formal dialectic". A fallacy would then be a violation of the rules of a formal dialogue game. This conception of a fallacy has been adopted in various formulations by Van Eemeren and Grootendorst [1984; 1992; 2004], Hintikka [1987] and Walton (e.g. [1998]). Walton has noted that fallacies often involve an illicit shift from one type of dialogue to another, typically into a quarrel where "anything goes". Krabbe has extended the consideration of a fallacy as a violation of the rules of a type of dialogue in which interlocutors are engaged, by developing "profiles of dialogue" that indicate among other things the way in which charges of committing a fallacy can be advanced and responded to [Krabbe, 1992; 1999]

How are fallacies to be classified? If a fallacy is a mistake in argumentation of a certain type, presumably one's taxonomy of fallacies will correspond to one's criteria for a good argument; for example if the individually necessary and jointly sufficient conditions for a good argument are acceptability of each premiss, relevance of each premiss to the conclusion drawn from it and sufficiency of the premisses in combination to support the conclusion [Johnson and Blair, 1993;

Freeman, 1991; Govier, 2005], then there will be three main types of fallacies: fallacies of unacceptability, fallacies of irrelevance, and fallacies of insufficiency.

How are individual fallacies to be analysed? In particular, are argumentative moves that have traditionally been stigmatized as fallacious sometimes legitimate? If so, under what circumstances? In a series of papers published between 1972 and 1982, and collected in [Woods and Walton, 1989], John Woods and Douglas Walton used the tools of formal logics other than classical first-order logic to explain a number of argumentative moves traditionally thought to be always fallacious (*argumentum ad verecundiam*, *petitio principii*, *argumentum ad baculum*, *argumentum ad hominem*, composition, division, *post hoc ergo propter hoc*, *ad ignorantiam*, *argumentum as populum*, equivocation, many questions) and to work out under what circumstances if any they in fact amounted to a fallacy. More recently, in a series of monographs and articles far too numerous to cite in full, Walton has used the approach of argumentation schemes and critical questions to distinguish legitimate from fallacious occurrences of such moves as arguing in a circle, appealing to popularity, and arguing against the person (see for example [Walton, 1998]).

A fine selection of contemporary work on fallacies, along with some classic historical papers, can be found in Hansen and Pinto [1995].

4. On argument criticism

What principles should govern the expression of one's evaluation of an argument in the form of argument criticism? Johnson [2000, 217–248] has proposed and argued for the following principles:

Principle of vulnerability: To be legitimate, an argument must be vulnerable to criticism.

Principle of logical neutrality: The critic should be clear about the nature of the criticism and should not pass off substantive criticism as logical criticism.

Principle of parity: Any line of reasoning or argument that is legitimate for one party to use is legitimate for the other.

Principle of discrimination: Criticisms of an argument should be balanced, kept in perspective, and integrated. Balance requires assessment of both strengths and weaknesses. Perspective requires that the discussion of an argument's problems focuses on the most important problems. Integration requires giving greatest emphasis to major criticisms.

5. On argument construction

What principles should govern argument construction?

What practical advice can be given for constructing good arguments?

ACKNOWLEDGEMENTS

I am deeply grateful for comments on an earlier version of this chapter to Lilian Bermejo-Luque, J. Anthony Blair, Inga B. Dolinina, Frans H. van Eemeren, Robert H. Ennis, Bart Garssen, Dale Hample, Ralph H. Johnson, Mike Metcalfe, Leah E. Polcar, Harvey Siegel and John Woods. They saved me from some embarrassing errors, and made useful suggestions. The responsibility for errors that remain is mine. Nor would they all agree with everything asserted in this chapter.

BIBLIOGRAPHY

[Anselm, 1903] Anselm, St. *Proslogium; Monologium: An Appendix In Behalf Of The Fool By Gaunilo; And Cur Deus Homo.* Translated from the Latin by Sidney Norton Deane, BA, with an introduction, bibliography, and reprints of the opinions of leading philosophers and writers on the ontological argument. Chicago: Open Court, 1903.

[Aristotle, 1984] Aristotle. *The Complete Works of Aristotle*, 2 vols., the revised Oxford translation edited by Jonathan Barnes. Bollingen Series LXXI: 2. Princeton: Princeton University Press, 1984.

[Austin, 1975] J. Austin. *How to Do Things with Words*, 2nd edition edited by J. O. Urmson and Marina Sbisà. Cambridge, Mass.: Harvard University Press, 1975. First edition first published in 1962.

[Beardsley, 1950] M. C. Beardsley. *Practical Logic*. Englewood Cliffs, NJ: Prentice-Hall, 1950.

[Blair, 2004] J. A. Blair. Argument and its uses. *Informal Logic* 24, 137–151, 2004.

[Blair and Johnson, 1987] J. A. Blair and R. H. Johnson. The current state of informal logic. *Informal Logic* 9, 147-151, 1987.

[Bolzano, 1985–89] B. Bolzano. *Wissenschaftslehre*, ed. Jan Berg. In his *Schriften* (Stuttgart-Bad Cannstatt: Frommann), vols. 11-13, 1985–1989. First published in 1837.

[Diogenes, 1925] Diogenes. *Lives of Eminent Philosophers*, with an English translation by R. D. Hicks. 2 vols. Loeb Classical Library numbers 184 and 185. Cambridge, Mass.: Harvard University Press, 1925. First published in the 3rd century CE (c. 210-240 CE).

[Eemeren and Grootendorst, 1984] F. H. van Eemeren and R. Grootendorst. *Speech Acts in Argumentative Discusssions: A Theoretical Model for the Analysis of Discussions Directed towards Solving Conflicts of Opinion.* Dordrecht-Holland / Cinnaminson-USA: Foris Publications, 1984.

[Eemeren and Grootendorst, 1992] F. H. van Eemeren and R. Grootendorst. *Argumentation, Communication and Fallacies: A Pragma-Dialectical Perspective.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.

[Eemeren and Grootendorst, 2004] F. H. van Eemeren and R. Grootendorst. *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach.* Cambridge: Cambridge Universsity Press, 2004.

[Ehninger and Brockriede, 1963] D. Ehninger and W. Brockriede. *Decision by Debate*. New York: Dodd, Mead, 1963.

[Ennis, 1982] R. H. Ennis. Identifying implicit assumptions. *Synthese* 51, 61-86, 1982.

[Euclid, 1969–73] Euclid. *Elementa*, vols. 14, ed. E. S. Stamatis (post J.L. Heiberg), 2nd edition. Leipzig: Teubner, 1969–1973.. First published ca. 300 BCE. Accessed via the *Thesaurus Linguae Graecae.*

[Feldman, 1994] R. Feldman. Good arguments. In Frederick F Schmitt (ed.), *Socializing Epistemology: The Social Dimensions of Knowledge*, Lanham, MD, Rowman and Littlefield, pp. 159–188, 1994.

[Finocchiaro, 1980] M. A. Finocchiaro. *Galileo and the Art of Reasoning: Rhetorical Foundations of Logic and Scientific Method.* Boston: Reidel, 1980.

[Freeman, 1991] J. B. Freeman. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure.* Berlin: Foris, 1991.

[Galilei, 1967] Galilei, Galileo. *Dialogue Concerning the Two Chief World Systems, Ptolemaic & Copernican*, 2nd. edition, translated by Stillman Drake, foreword by Albert Einstein. Berkeley: University of California Press. 1967.

[Gelder, 2004] T. J. van Gelder. *Reason!Able*, 2004. Available at `http://www.goreason.com/`. Visited 3 July 2005.

[George, 1983] R. George. Bolzano's consequence, relevance, and enthymemes. *Journal of Philosophical Logic* 12, 299-318, 1983.

[Gilbert, 1997] M. A. Gilbert. *Coalescent Argumentation.* Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

[Goldman, 1999] A. I. Goldman. *Knowledge in a Social World.* New York: Oxford University Press, 1999.

[Govier, 1987] T. Govier. *Problems in Argument Analysis and Evaluation.* Studies of Argumentation in Pragmatics and Discourse Analysis, 5. Dordrecht, Holland: Mouton de Gruyter/Foris, 1987.

[Govier, 2005] T. Govier. *A Practical Study of Argument*, 6th edition. Belmont, CA: Wadsworth, 2005. Previous editions published 1985, 1988, 1992, 1997, 2001.

[Grennan, 1997] W. Grennan. *Informal Logic: Issues and Approaches.* Montreal & Kingston: McGill-Queen's University Press, 1997.

[Groarke, 1996] L. Groarke. Logic, art and argument. *Informal Logic* 18, 116-131, 1996.

[Hamblin, 1970] C. L. Hamblin. *Fallacies.* New York: Methuen, 1970.

[Hansen and Pinto, 1995] H. V. Hansen and R. C. Pinto. *Fallacies: Classical and Contemporary Readings.* University Park, PA: Pennsylvania State University Press, 1995.

[Hastings, 1963] A. C. Hastings. *A Reformulation of the Modes of Reasoning in Argumentation*, Ph.D. dissertation, Northwestern University, 1963.

[Hintikka, 1987] J. Hintikka. The fallacy of fallacies. *Argumentation* 1, 211-238, 1987.

[Hitchcock, 2002] D. Hitchcock. The practice of argumentative discussion. *Argumentation* 16, 287–298, 2002.

[Johnson, 2000] R. H. Johnson. *Manifest Rationality: A Pragmatic Theory of Argument.* Mahnwah, NJ: Lawrence Erlbaum Associates, 2000.

[Johnson, 2002] R. H. Johnson. Manifest rationality reconsidered: reply to my fellow symposiasts. *Argumentation* 16, 311-331, 2002.

[Johnson and Blair, 1993] R. H. Johnson and J. A. Blair. *Logical Self-Defense*, 3rd edition. Toronto: McGraw-Hill Ryerson, 1993. First edition published 1977, 2nd edition 1983, 1st U.S. edition 1994.

[Johnson and Blair, 2000] R. H. Johnson and J. A. Blair. Informal logic: an overview. *Informal Logic* 20, 93-107, 2000.

[Kienpointner, 1992] M. Kienpointner. *Alltagslogik: Struktur und Funktion von Argumentationsmustern.* Stuttgart: FrommannHolzbog, 1992.

[Kneale and Kneale, 1962] W. C. Kneale and M. Kneale. *The Development of Logic.* Oxford: Clarendon Press, 1962.

[Krabbe, 1992] E. C. W. Krabbe. So what? Profiles for relevance criticism in persuasion dialogues. *Argumentation* 6, 271–283, 1992.

[Krabbe, 1999] E. C. W. Krabbe. Profiles of dialogue. In Jelle Gerbrandy, Maarten Marx, Maarten de Rijke, and Yde Venema (eds.), *JFAK: Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*, III, Amsterdam: Amsterdam University Press, pp. 25–36, 1999.

[Perelman and Olbrechts-Tyteca, 1958] C. Perelman and L. Olbrechts-Tyteca. *La nouvelle rhétorique. Traité de l'argumentation.* Bruxelles: Editions de l'Université de Bruxelles, 1958.

[Perelman and Olbrechts-Tyteca, 1969] C. Perelman and L. OlbrechtsTyteca. *The New Rhetoric: A Treatise on Argumentation*, trans. John Wilkinson and Purcell Weaver. Notre Dame: University of Notre Dame Press, 1969.

[Pinto, 2001] R. C. Pinto. *Argument, Inference and Dialectic: Collected Papers on Informal Logic with an Introduction by Hans V. Hansen.* Dordrecht / Boston / London: Kluwer, 2001.

[Pinto, 2001a] R. C. Pinto. Generalizing the notion of argument. In Pinto 2001, pp. 10-20. First published in Frans H. van Eemeren, Rob Grootendorst, J. Anthony Blair and Charles Willard (eds.), *Proceedings of the Second International Conference on Argumentation*, vol. 1A (Amsterdam: SICSAT, 1991), pp. 137-146, 2001.

[Pinto, 2001b] R. C. Pinto. The relation of argument to inference. In Pinto 2001, pp. 32-45. First published in Frans H. van Eemeren, Rob Grootendorst, J. Anthony Blair and Charles Willard (eds.), *Perspectives and Approaches: Proceedings of the Third International Conference on Argumentation*, vol. 1 (Amsterdam: SICSAT, 1995), pp. 271-286. 2001.

[Plantin, 2003] C. Plantin. Argumentation studies in France: a new legitimacy. In Frans H. van Eemeren, J. Anthony Blair, Charles A. Willard and A. Francisca Snoeck Henkemans (eds.), *Anyone Who Has a View: Theoretical Contributions to the Study of Argument* (Dordrecht / Boston / London: Kluwer), pp. 173–187, 2003.

[Plato, 1997] Plato. *Complete Works*, edited, with introduction and notes, by John Cooper, associate editor, D. S. Hutchinson. Indianapolis: Hackett, 1997.

[Reed and Rowe, 2005] C. Reed and G. Rowe. *Araucaria Version 3.0*, 2005. Available at `http://aravcaria.computing.dundee.ac.uk/`. Visited 3 July 2005.

[Rolf and Magnusson, 2002] B. Rolf and C. Magnusson. . *Athena Standard 2.5*, 2002. Available at `http://www.athenasoft.org/sub/software.htm`. Visited 3 July 2005.

[Searle, 1969] J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. London: Cambridge University Press, 1969.

[Searle, 1979] J. R. Searle. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press, 1979.

[Sextus Empiricus, 1935] Sextus Empiricus. *Against the Logicians*, trans. R. G. Bury. Cambridge, MA: Harvard University Press, 1935. First published c. 180-200 CE.

[Siegel and Biro, 1997] H. Siegel and J. Biro. Epistemic normativity, argumentation, and fallacies. *Argumentation* 11, 277-292, 1997.

[Snoeck Henkmans, 1992] A. F. Snoeck Henkemans. *Analyzing Complex Argumentation*. PhD dissertation, University of Amsterdam, 1992.

[Thomas, 1973] S. N. Thomas. *Practical Reasoning in Natural Language*, 1st edition. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[Toulmin, 1958] S. Toulmin. *The Uses of Argument*. Cambridge: Cambridge University Press, 1958.

[Vorobej, 1994] M. Vorobej. The TRUE test of linkage. *Informal Logic* 16, 147-157, 1994.

[Vorobej, 1995] M. Vorobej. Hybrid arguments. *Informal Logic* 17, 289–296, 1995.

[Walton, 1990] D. N. Walton. What is reasoning? What is an argument? *Journal of Philosophy* 87, 399-419, 1990.

[Walton, 1996a] D. N. Walton. *Argument Structure: A Pragmatic Theory*. Toronto: University of Toronto Press, 1996.

[Walton, 1996b] D. N. Walton. *Argumentation Schemes for Presumptive Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

[Walton, 1998] D. N. Walton. *Ad Hominem Arguments*. Tuscaloosa and London: University of Alabama Press, 1998.

[Walton and Krabbe, 1995] D. N. Walton E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Albany, NY: State University of New York Press, 1995.

[Wellman, 1971] C. Wellman. *Challenge and Response: Justification in Ethics*. Carbondale, IL: Southern Illinois University Press, 1971.

[Wenzel, 1990] J. W. Wenzel. Three perspectives on argument: Rhetoric, dialectic, logic. In J. Schuetz & R. Trapp (eds.), *Perspectives on argumentation: Essays in honor of Wayne Brockriede* (Prospect Heights, IL: Waveland), 9–26, 1990.

[Whately, 1848] R. Whately. *Elements of Logic*, 9th edition. London: Longmans, 1848. First edition published 1826.

[Woods and Walton, 1989] J. Woods and D. N. Walton. *Fallacies: Selected Papers 1972-1982*. Dordrecht, Holland and Providence, RI: Foris, 1989.

ON THE RELATION OF INFORMAL TO SYMBOLIC LOGIC

Dale Jacquette

The distinction between formal and informal logic is clarified as a prelude to considering their ideal relation. Aristotle's syllogistic describes forms of valid inference, and is in that sense a formal logic. The square of opposition and rules of middle term distribution of positive or negative propositions in an argument's premises and conclusion, in contrast, are generally understood as devices of so-called informal logic and critical reasoning. I propose a more exact criterion for distinguishing between formal and informal logic, and then defend a model for fruitful interaction between informal and symbolic methods of investigating and critically assessing the logic of arguments.

1 A STRANGE DICHOTOMY

In the history of logic a division between formal and informal methods has emerged. The distinction is not always sharply drawn, and there are theoretical disagreements with practical implications about the logical techniques that should be classified as formal or informal, and about the scope and limits and paradigms of each of these two main categories of logic.

The distinction arises because of unresolved questions about how logical relations are best expressed. At its furthest extremes, the distinction manifests itself as a difference of ideology and methodology between practitioners of the most highly formalized symbolic mathematical logics and the purely informal treatment of logical inference in the rhetoric of argumentation, using primarily the discursive resources of ordinary language while avoiding formal logical or mathematical symbolisms.

The polemics associated with disputes between formal and informal logicians are instructive even when they are not especially worthy of imitation. Formal logicians of a radical stripe often dismiss informal logical techniques as insufficiently rigorous, precise, or general in scope, while their equally vehement counterparts in the informal logic camp typically regard algebraic logic and set theoretical semantics as nothing more than an empty formalism lacking both theoretical significance and practical application when not informed by the informal logical content that formal logicians pretend to despise. Anyone who has spent time working in formal or informal logic, and certainly those who have crossed over from time to time to try their hand at both styles of logic or associated with outspoken members of

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

each side should have direct anecdotal experience of the polarization among some formal and informal logicians.

I do not insist that the antagonism between formal and informal logicians cannot be friendly and respectful, although I suspect that unfortunately it is often something less. There is no reason for example why a healthy intellectual rivalry between formal and informal logicians should not be fostered. A properly managed division of labor might promote advances on both sides in the sort of free market competition of ideas that can help logic to flourish. Such is evidently the case with respect to the tolerance among logicians and mathematicians of differences between classical and intuitionistic and other nonstandard logics, Euclidean and non-Euclidean geometry, and even more prosaic distinctions of method such as that between axiomatic and natural deduction techniques in symbolic logic. The impression is inescapable, however, that the current situation in logic with its deep-seated bifurcation between formal and informal methods is in this regard far less than ideal. The division is something more like that described by C.P. Snow in his landmark discussion of *The Two Cultures.*[1] The analogy is appropriate in so many ways that the distinction between formal and informal logic may turn out to be just a special instance of the bifurcation of science and the humanities described by Snow, with formal logicians falling on the same side as the exact sciences, and informal logic finding its natural home among the humanities.

Now I am struck by the division between formal and informal logic as constituting a strange dichotomy. Since I consider myself to be both a formal and informal logician, I see the formal and informal logical extremes as two complementary poles along a continuum that insensibly grade off into one another without a definite boundary, and hence as offering no clearcut basis for partisanship. I agree with the informal logicians' claim that symbolic logic by itself is an empty formalism that needs to have a content supplied by informal reasoning. But I also agree with formal logicians when they complain that purely informal logic is often too blunt an instrument to handle sophisticated problems involving complex logical relations, and that informal reasoning can sometimes overlook logical problems and solutions that formal logic is uniquely able to reveal. I am inclined on pragmatic grounds, therefore, to accept whatever logical methods are best suited for my analytic purposes in trying to understand different types of logical problems. I think it is only reasonable, as a result, not to stand on any exclusionary principle in confronting the logic of arguments, but to adopt a style of doing logic that appears to work best in formulating ideas and testing inferences, in expressing and critically analyzing concepts and arguments, and generally speaking as a guide to good thinking.

I assume that just as there are adversaries who gravitate toward one side of the formal-informal logic distinction in dialectical opposition to the other, so there must also be many others like myself who prefer a principled compromise between the two extremes, who hope that formal and informal logic can work together in a productive partnership. It is to give substance to this potential for integrating formal and informal logic that I propose first to offer a more exact criterion of the distinction between formal and informal logic, and then to defend a model for the fruitful interaction between formal and informal logical methods.

---

[1] [Snow, 1964].

2 TOWARD A DEFINITION OF FORMAL AND INFORMAL LOGIC

If the attempt to find middle ground between exclusionary formal and informal logic programs is to succeed, then the first task is to clarify the concepts of formal and informal logic themselves. Then it may be possible to describe a way of reconciling formal and informal logical techniques in a more comprehensive philosophy of logic.

The concepts of formal and informal logic are best approached by beginning with some faulty attempts at definition, and trying to understand exactly how and why they go wrong. We have already observed that Aristotle's logic of syllogisms with its rules for valid inferences and square of opposition is a kind of formal logic, even though it is usually classified as a branch of informal logic.

The same is true of Beardsley argument diagramming and Venn, Euler, and Lewis Carroll diagramming. These methods of representing the structure and assessing the deductive validity especially of syllogistic arguments involving Boolean relations as set theoretical inclusions or exclusions of entities in the extensions of predicates intuitively concern the formal logical properties of propositions and arguments. The use of schematic variables, mathematically regimented syntax, and graphic or set theoretical semantic techniques, moreover, are not necessary to distinguish formal from informal logics. A formal logic in one sense can always be presented in ordinary albeit stinted English. The connections are explicit at least in simple cases. We see this clearly in the effort to formulate in symbolic logic an expression such as $(\exists x)(F = \lambda y[\Box(y = x)])$, and in articulating in nonsymbolic but still formal terms the equivalent proposition that 'There exists an entity such that a certain property is identical to the abstract property of being necessarily identical to the entity'. In the English paraphrase there is no 'formalism' or symbolic notation as usually understood; yet the phrase as much describes formal relations as does the more straightforwardly 'formal' symbolism containing the existential quantifier, identity sign, lambda operator, modal necessity operator, parentheses and brackets. All such elements are legitimately part of formal logic insofar as they involve logical forms of relations and logical forms of arguments, regardless of whether they are written in a special symbolic notation or in modified English.

It is tempting in light of these conclusions to consider eliminating the distinction between formal and informal logic altogether. This is an attractive option if we interpret all of logic as involving the forms and structures of syntax, semantic relations, and argumentive patterns. Then, while the term 'logical form' remains to designate a special kind of form, the phrase 'formal logic' is technically redundant. It may then be preferable to substitute the terminology of symbolic-nonsymbolic logic for the formal-informal logic distinction. The disadvantage of such an approach is that the distinction between formal and informal logic is well-entrenched in logicians' parlance. Arguably, we do less violence to established usage if in good linguistic conscience we maintain the formal-informal logic vocabulary by providing acceptable definitions of 'formal' and 'informal' logic. The trouble is that I am not at all sure that there can be a satisfactory definition of the distinction between formal and informal logic that is both intuitively correct and fully in agreement with current terminology. The reason for this skepticism is that established linguistic practice among logicians is itself not entirely correct. I shall accordingly try to negotiate a commonsense compromise that distinguishes between formal and informal logic in a way that does not exactly conform to the standard convention, but that I think many and perhaps even most logicians on both sides of the distinction may find acceptable.

My suggestion is to distinguish between formal and informal logic by applying the arguably less controversial because criteriologically better-defined distinction between specialized symbolic versus nonspecialized nonsymbolic logic. More specifically, I propose that a logical theory or procedure is formal if and only if it adopts a specialized symbolism for representing logical forms that does not occur in ordinary nonspecialized nonsymbolic thought and language. Although I acknowledge that all of logic has to do with logical form, I do not agree that all expressions of logical form must themselves be formal. This distinction captures much of the received concept, since it includes all of symbolic logic and excludes nonsymbolic evaluations of validity or invalidity. As we might expect, formal logic by the proposed distinction will roughly include everything belonging to what has become the *de facto* criterion for formal logic in relations expressed by means of standard and nonstandard notational variations and extensions of the propositional and predicate-quantificational calculus. The definition additionally includes schematic and graphic treatments of syllogistic logic that have traditionally been regarded as more properly within the aegis of informal logic and critical reasoning. Informal logic by contrast on the present proposal is limited to the consideration of a proposition's or argument's logical form by discursive reconstruction within natural language, the use of counterexamples to discredit inferences, identification of arguments as committing any of the so-called rhetorical fallacies, and the like.

The relegation of syllogistic logic, square of opposition, and argument diagramming methods to the genus of informal logic can now be seen as a kind of historical accident. Were it not for the emergence of more powerful algebraic methods of formal logic with the discovery of formal logical techniques in Frege's *Begriffsschrift* and C.S. Peirce's proto-quantificational logic, there is little doubt that the logic of syllogisms, Venn and other styles of diagramming, etc., would constitute the whole of formal logic as opposed to purely informal nonspecialized nonsymbolic logical methods. Logic would then still be today as if it had never advanced beyond the point concerning which Kant in the late eighteenth century was able to declare that no significant advances had been made since the time of Aristotle.[2] Why, then, should such logical techniques be displaced as informal given the development of contemporary algebraic and rigorously algorithmic methods of mathematical logic?

---

[2] [Kant, 1965, Bviii]: "That logic has already, from the earliest times, proceeded upon this sure path is evidenced by the fact that since Aristotle it has not required to retrace a single step,

On this proposal, it is more appropriate to classify syllogisms and the tools of logic that have standardly been turned over to the informal logic and critical reasoning textbooks as less powerful, general, and technically advanced, but every bit as formal as symbolic mathematical logics. As a consequence, I include Aristotelian syllogistic and all related graphic and diagramming paraphernalia as part of genuinely formal logic. I am therefore committed to saying that these methods are properly part of formal logic despite their usually being included in what are called informal logic texts as adjuncts to what is standardly presented as the informal logic curriculum. If this is true, then it may be time for logicians to admit that insofar as they use syllogistic logic and argument diagramming they are doing formal logic, frequently under the mistaken rubric of informal logic. It may equally be time in that case for formal logicians to admit that there are weaker less universal methods of logic that are just as formal as the algebraic methods of formal symbolic logic, which they may prefer to use, but which do not for that reason alone have exclusive title to the category of formal logic.

What, then, is to be said about the English translation of $(\exists x)(F = \lambda y[\Box(y = x)])$? The natural language equivalents of any such expression are strictly informal by my criterion. Yet I imagine that ordinary language translations of expressions in symbolic logic are unlikely to find application in practice as a workable replacement for mathematical logic. The reason quite obviously is that the ordinary language equivalents while available in principle are simply too cumbersome, as they would be in the parallel case of trying to develop an informal mathematics for the purpose of formulating and solving quadratic polynomial equations. With regard to another problem of the practical implications of the distinction, I have no definite pedagogical recommendations to make with respect to the division of logical methods according to the traditional distinction between formal and informal logic. Thus, I see nothing wrong with continuing to teach syllogistic logic and Venn diagramming, to mention two conspicuous examples, as part of nonmathematical logic, critical reasoning, or what is popularly called informal logic. It is just that by my characterization of the distinction between formal and informal logic I find it misleading to refer to such techniques as informal.

I do not claim that my definition of formal and informal logic could not be improved. I have already indicated that I regard the proposal as a tradeoff between conflicting jointly unsatisfiable desiderata. For present purposes, I want to explore the proposed criterion by investigating the implications of distinguishing between formal and purely informal logic alternatively as the analysis of logical form by means of specialized symbolisms versus the study of logical form by means of ordinary language.

---

unless, indeed, we care to count as improvements the removal of certain needless subtleties or the clearer exposition of its recognised teaching, features which concern the elegance rather than the certainty of the science. It is remarkable also that to the present day this logic has not been able to advance a single step, and is thus to all appearance a closed and complete body of doctrine."

3 FORMAL AND INFORMAL LOGIC IN PARTNERSHIP

There are many ways in which formal and informal logic interact. There are situations in which formal and informal logic can profitably cooperate in order to critically evaluate arguments.

Formal symbolic logics are always accompanied by and presented within a discursive framework of informal metalanguage introductions and explanations, or can be traced back through a genealogy of formal conventions to an informal context. Without grounding in ordinary language and a relation to informal ideas, even the formalisms most familiar to practicing logicians lack meaning and application. If symbolic logic is not always needed, if it can be an impediment to understanding, and if it cannot function effectively entirely on its own for theoretical purposes in the explication of logical connections and deductive proof of consequences, then the use of specialized formalisms must be justified by a philosophical rationale. Informal logic is useful and often essential in working through a preliminary heuristic analysis of a problem before it can be decided whether and if so what kind of formal logic to apply in modeling a given choice of logical relations or in solving a logical problem. Sometimes informal methods provide a better, easier, or more understandable conceptual analysis of the logic of a proposition or argument.

Accordingly, I want to propose a pragmatic principle that allows informal and formal logical methods to be used individually or in combination to achieve the best analysis of the logic of arguments as determined by the specific requirements of each situation. The ideal is for logicians to cultivate proficiency in as many formal and informal logical methods as may be available, not excluding efforts to discover or invent new techniques as each task may demand. The exigencies of each analytic problem should be considered independently on its own terms as a challenge for logical investigation in its own right. I illustrate the potential gains from this partnership between formal and informal logic with two examples. The first application demonstrates the importance of informal reasoning as a preparation for formalization in the informal analysis of logical problems. The preliminary informal consideration of a logical paradox when properly done can sometimes avoid the need for formalization altogether, while in other cases it may demarcate the extent to which a minimal partial formalization may be useful. Alternatively, informal evaluation can provide a set of parameters by which the crucial concepts and relations to be formalized can first be identified and an exact purpose and justification for formalization determined. The problem I have chosen is the Epimenides or liar paradox, a famous logical puzzle with interesting implications for our understanding of the nature of truth, the limits of sentence meaning in a language, and the semantics of reference, for which many different elaborate formal logical solutions have been suggested.

The liar paradox

The liar paradox is the implicit logical inconsistency entailed by a sentence that asserts its own falsehood. There are several different formulations of liar sentences, which for convenience can be represented by a simplified expression of sentential self-denial:

(L) This sentence is false.

Assuming a classical bivalent truth value semantics, the liar paradox is supposed to follow by dilemma from the constructibility of L. This is the orthodox informal explanation of the liar paradox dilemma. Suppose L is true. Since L says of itself that it is false, then, if L is true, it is false. Now suppose that L is false. Then, again because of L's self-denial, it is not the case that L is false, from which it follows that L is true. The conclusion seems inescapable that liar sentence L is true if and only if it is false.[3] The contradiction is intolerable in classical bivalent propositional logic and semantics. There are two main categories of responses to the liar paradox within standard logic. The force of the liar paradox is often understood to exert pressure on the concept of truth and the bivalent truth value semantics in which the paradox arises. Alternatively, the truth-valuability or meaningfulness of the liar sentence or its constructibility in an ideal language is challenged.

Many solutions have been proposed. Historically, these have included introducing nonstandard many-valued or truth-value gap logics to replace the ordinary two-valued structure, and legislating against self-reference, or self-(non)-application of indexicals, demonstratives, or truth value predications within a language. The former strategy has been shown by Brian Skyrms in "Return of the Liar: Three-Valued Logic and the Concept of Truth" to be ineffective by itself against strengthened liar paradoxes that can be constructed for any enlargement of truth-value semantics. The idea is to define a new truth function that collapses $n \geq 2$ truth values into two, by mapping onto falsehood whatever truth value negation receives in the n-valued system.[4] The latter strategy is at the heart of Alfred Tarski's influential theory of truth, in which formal languages are stratified into object- and

---

[3] An excellent introduction is [Martin, 1970]. See also [Barwise and Etchemendy, 1987; Martin, 1984]. The most complete historical discussion of the Epimenides paradox (Paul, Epistle to Titus 1 verses 12-13) is [Rüstow, 1987].

[4] See [Skyrms, 1970, 153–161]. If we try to forestall the liar paradox dilemma by introducing a third truth value and assigning it to the liar sentence, then we can define a new truth table and reformulate the liar paradox in this way. For nonstandard trivalent negation with a third undetermined (U) truth value, we might have:

p	$\neg p$
T	F
F	T
U	U

Now we define:

p	$@p$
T	F
F	T
U	F

metalanguages, and truth value predications are forbidden within any given language in the hierarchy. Truth value predications can then only be made about a lower-level in a higher-level metalanguage in an indefinitely ascending order of object- and metalanguages in which it is possible to make meaningful pronouncements only about the semantic properties of lower-level languages.[5] Saul A. Kripke's "Outline of a Theory of Truth" combines both truth-value gaps and a modified Tarskian stratification of object- and metalanguages, permitting transfinitely ramified occurrences of metalanguages within each level of the original Tarskian hierarchy. Kripke diagnoses the logical ailment of the liar sentence as a failure of 'grounding', since the sentence seems indefinitely to flip-flop between truth and falsehood without reaching any settled stable state representing a single fixed semantic value.[6]

What is common to a certain widely shared response to the liar sentence is the assumption that the mere constructibility of the liar sentence is sufficient to entail the liar paradox — that the liar sentence is true if and only if it is false. The inferability as opposed to the mere constructibility of the liar sentence would be quite another thing. No one, however, has attempted to prove that the liar sentence can be deduced from an otherwise logically consistent set of sentences, such as a scientific or philosophical theory. Logicians and philosophers who fret about the liar sentence often assume that the only way to avoid the paradox is by preventing the construction of the liar sentence. The attitude is typically that, unless appropriate precautions are taken, the liar sentence can be formulated, after which the liar paradox inevitably follows. To forestall outright logical contradiction, we must either prevent its construction or take steps, sometimes amounting to the abandonment of classical propositional logic, to ameliorate its disastrous effects. The liar sentence exists or can be uttered, so we had better do something about it. It is precisely this otherwise reasonable conventional assumption that I now, informally, want to question.[7]

If the mere constructibility of the liar sentence in a language entails the liar paradox, then the following dilemma inference must hold for some sentence S:

The strengthened liar paradox then presumably holds when we construct a strengthened liar sentence L^* that says '@(This sentence is true)' (instead of this equivalent of the standard formulation, '$\neg$(This sentence is true)'), and then subject the strengthened liar sentence L^* to a modified dilemma that takes as its basis the disjunction $L^* \vee @L^*$ instead of $L \vee \neg L$. The method is generalizable to any expanded many-valued truth value matrix involving any number of truth values.

[5] [Tarski, 1956, 152–278].

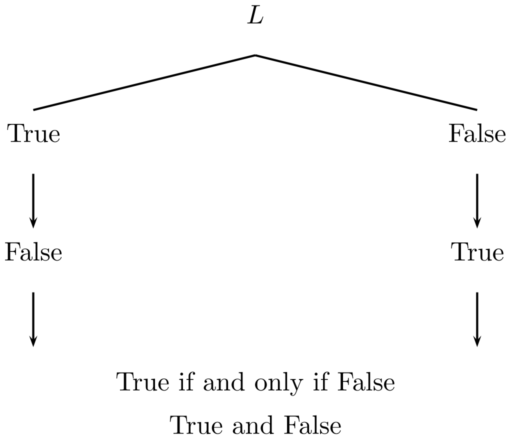
[6] [Kripke, 1975, 690–716].

[7] A typical statement of the orthodox account by which the constructibility of the liar sentence alone entails the liar paradox appears in [Tarski, 1956, 158]: "The source of this contradiction is easily revealed: in order to construct the assertion (β) ['*c is not a true sentence*' *is a true sentence if and only if c is not a true sentence*] we have substituted for the symbol 'p' in the scheme (2) [*x is a true sentence if and only if p*] an expression which itself contains the term 'true sentence'... Nevertheless no rational ground can be given why such substitutions should be forbidden in principle."

$$(D)\quad \frac{L \lor \neg L}{S \land \neg S}$$

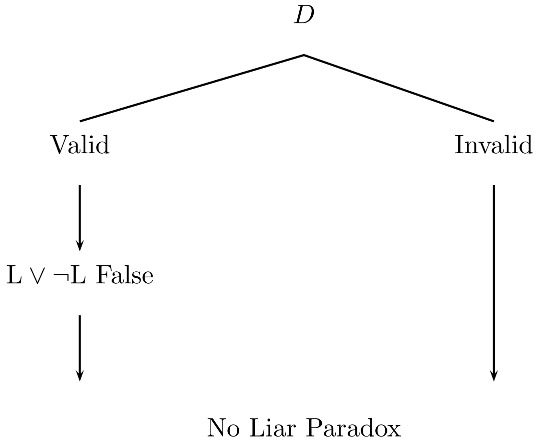
Within the standard semantic framework in which the liar paradox and its solutions are considered, the inference presumably involves a valid deductive inference from a logically consistent, even logically necessary, sentence to another self-contradictory sentence. If the liar sentence L is constructible, then a sentence of the form $S \land \neg S$, such as 'L is true and it is not the case that L is true' (that is, where $S = L$), is thought paradoxically to be logically forthcoming in the dilemma by excluded middle from the disjunctive tautology $L \lor \neg L$, given the self-truth-denying content of L.

Another way to express the dilemma as follows:



We now need to ask whether the dilemma inference D is itself logically valid or invalid. There is a dilemma that proves a semantic metaparadox about the liar paradox from the same classical semantic assumption that D is either valid or invalid. Let us consider the possibilities in opposite order. Suppose that D is deductively invalid. Then the liar paradox is not entailed by the mere constructibility or truth or falsehood of liar sentence L. For, in that case, the liar sentence whether true or false by hypothesis does not logically imply an explicit syntactical contradiction of the form $S \land \neg S$ for any sentence S. If there is to be a liar paradox engendered by the constructibility of the liar sentence L, then dilemma inference D must on the contrary be deductively valid. If D is valid, however, then, since the conclusion of the inference $S \land \neg S$ is standardly false, and, indeed, deductively invalid, it follows that the liar sentence dilemma based on the disjunction $L \lor \neg L$ itself must be logically necessarily false. This is a rather extraordinary, but, I think, unavoidable conclusion, which to my knowledge has nowhere been acknowledged in the extensive philosophical literature on the liar. If the liar sentence dilemma or excluded middle involving the liar sentence cannot be true in order to entail the liar paradox, then there simply is no liar paradox even on the assumption that the liar sentence *L* is true. The reason is that the liar sentence entails the liar paradox by dilemma inference *D* only if the liar sentence is both true and false, or such that it is true if and only if it is false. We might not be satisfied to have the liar sentence turn out to be necessarily false in this way, but our discomfort cannot be understood as a result of the entailment of a logical paradox from the mere constructibility of the liar sentence.

We can represent the liar metaparadox by a similar dilemma structure as above for the liar paradox:



We are not yet out of the woods. The metaparadox I have informally described is not offered as in any sense a 'solution' to the liar paradox. We cannot conclude on the basis of the metaparadox that the liar sentence is simply logically necessarily false. For without special provision, that seems to just put us back on the standardly semantically problematic false → true, true → false loop. If a sentence expressing its own falsehood is false or even and especially if it is logically necessarily false, then, apparently, it is true. This is no progress over the usual presentation of the liar paradox. The point I am trying to make is not that the liar sentence is false, but rather that the liar paradox is not standardly entailed by the mere constructibility of the liar sentence. If the liar paradox is not entailed by the mere constructibility as opposed to the inferability of the liar sentence, then the mere constructibility of the liar sentence poses no threat to classical bivalent truth value or validity semantics, nor to the naive folk theory of truth. Hence, the mere constructibility of the liar sentence in a language provides no adequate philosophical motivation for the elaboration of a special formal semantic policy or revision of the naive theory, such as that represented by a semantics of truth value gaps or Tarskian hierarchy of object- and metalanguage truth value predications, or Kripkean hybrid of these earlier proposals.

There is a semantic confusion that remains to be untangled. We must recall the over-arching assumption in which the necessary falsehood of the liar sentence is considered. It has not been suggested that the liar sentence L is false *simpliciter*, but only that the liar is standardly logically necessarily false *if* or *on the assumption that* dilemma inference D is standardly logically valid. If L is true if and only if it is false if or on the assumption that dilemma inference D is classically deductively logically valid, then there is only one possible conclusion, which is that dilemma inference D is classically deductively *in*valid. We can, and, as a matter of fact, we must proceed in this way. If we do, however, then we are back in the first liar metaparadox horn, according to which, where dilemma inference D is logically invalid, the liar paradox once again is not entailed by the mere constructibility of the liar sentence. If dilemma inference D is logically valid, then liar sentence L is logically necessarily false, from which it seems to follow that L, which ostensibly says of itself that it is false, is also logically necessarily true. This conclusion again holds not in any absolute sense, but only conditionally on the assumption that dilemma inference D is logically valid. What follows therefore is rather that dilemma inference D is not logically valid, and that the liar paradox is not entailed by the mere constructibility of the liar sentence.

The semantic metaparadox about the liar paradox shows that whether the liar dilemma is deductively valid or invalid, the liar paradox is not logically entailed by the mere constructibility of the liar sentence. The dissolution of the liar paradox via the liar semantic metaparadox shows that the liar paradox cannot be intelligibly formulated as entailed by the mere constructibility of the liar sentence in languages powerful enough to express truth-self-denials. The liar metaparadox suggests that something is wrong with the usual description of how the liar sentence is supposed to entail the liar paradox by way of the liar paradox dilemma — but *what*? What hidden fallacy invalidates the liar paradox argument?

An interesting question remains unanswered if we grasp the invalidity horn of the liar metaparadox dilemma. Here we confront the problem of what makes the liar paradox dilemma in D deductively invalid. From an intuitive point of view, it seems inadequate to say merely that the inference is invalid because the liar sentence dilemma $L \vee \neg L$ as a tautology of propositional logic is necessarily true, while the conclusion $S \wedge \neg S$ as a contradiction or inconsistency of propositional logic is necessarily false. This naturally makes the paradox dilemma deductively invalid. A deeper difficulty nevertheless remains to challenge our understanding of the content of liar sentence L. It requires of any dissolution of the liar paradox by semantic metaparadox a reasonable explanation of why in particular it is deductively invalid in the second liar paradox dilemma horn to make use of either or both of the following inferences (or to validly detach the consequents from their respective material conditional counterparts): $L \vee \neg L$

(i) L is true $\vdash$ L is false (L is true $\rightarrow$ L is false) ($L \rightarrow \neg L$)

(ii) L is false $\vdash$ L is true (L is false $\rightarrow$ L is true) ($\neg L \rightarrow L$)

Why, we must ask, do these inferences not go through? Given the apparent content of sentence L as denying its own truth, why does the falsehood of sentence L not validly deductively follow from its truth, or its truth from its falsehood, as the colloquial description of the liar paradox seems to require?

It may be worthwhile at this juncture to take notice of some of the standard and nonstandard reactions to the content of the liar sentence in relation to the liar paradox. Some logicians have concluded that both the liar sentence L and its negation $\neg L$ are meaningless, and hence neither true nor false. This maneuver blocks the valid inferences or true material conditionals in (i) and (ii) standardly needed to derive the liar paradox from the constructibility of the liar sentence. Wittgenstein in the *Tractatus Logico-Philosophicus* dismisses the liar sentence as meaningless because of the picture theory of meaning. He argues that no meaningful sentence can 'get outside itself' or include within itself a picture of itself with the same logical form or mathematical multiplicity of one-one corresponding elements under analysis. This he claims any construction *per impossibile* must do in order to picture an existent or nonexistent fact about itself.[8] Tarski, in "The Concept of Truth in Formalized Languages", argues that the liar paradox motivates an indefinitely ascending hierarchy of object-languages and metalanguages to implement the formation principle by which no meaningful sentence in a formalized language can express the truth or falsehood of other sentences belonging to the same language.[9] More recently, paraconsistent logics that in different ways tolerate syntactical contradictions allow the liar sentence and its negation to be regarded as both true and false. This strategy accommodates the intuitive sense by which the liar paradox is supposed to follow from the content of the liar sentence in the orthodox informal characterization. Graham Priest, for example, in *Beyond the Limits of Thought* (without reference to the liar in this immediate context), maintains that: "…if α is both true and false, so is $\neg\alpha$, and so is $\alpha \wedge \neg\alpha$. Hence, a contradiction can be true (if false as well)."[10]

I do not strongly feel the need to refute any of these resolutions, each of which in my opinion has some merits and some disadvantages. Instead, I shall briefly remark some of the difficulties I find in each. The sustained criticism Wittgenstein offers of the picture theory of meaning in the first third of his *Philosophical Investigations* constitutes good grounds for preferring an alternative semantics.[11] I pass without further comment on the most trenchant criticisms frequently raised against Tarski's theory of truth. These are that by refusing to permit truth predications to be made of sentences within the same language Tarski throws out the baby with the bathwater, and that the indefinite hierarchy of object- and metalanguages for truth value predications never achieves a final characterization of the concept of truth. Tarski's metalinguistic restrictions on truth predications not only disallow diagonal constructions like the liar sentence, to the relief of those who are concerned about its mere constructibility entailing semantic paradox, but also all logically harmless truth value pronouncements, such as 'This sentence is true'. The indefinitely ascending stratification of metalanguages in which the truth or falsehood of sentences is permitted only for lower tiers of the hierarchy never reaches an endpoint at which the theorist can say that truth has finally been defined. The Tarskian semantic hierarchy furthermore is self-consciously *ad hoc*, supported by no independent justification other than its apparent usefulness in solving the liar paradox.[12] Kripke's solution mentioned earlier, involving truth value gaps in truth predications transfinitely iterated within each tier of a Tarskian structure of metalanguages seems unnecessarily complicated. The paraconsistent analysis of the liar, despite the appeal of its frank avowal of the liar sentence as dialethically both true and false, strikes me as bizarre. If the liar sentence is both true and false, then the liar dilemma inference to the liar paradox would appear to be both valid and invalid.[13]

---

[8] [Wittgenstein, 1922], 2.174; 3.332-3.333 (concerning Russell's theory of types). In 3.332 Wittgenstein writes: "No proposition can say anything about itself because the propositional sign cannot be contained in itself…", and in 4.442: "A proposition cannot possibly assert of itself that it is true."

[9] [Tarski, 1956, 157–165 and *passim*].

[10] [Priest, 1995, 156-161]. For indications of dialethic intutions about the proper response to the liar paradox, see also [Priest and Routley, 1989, 12-13, 22-23, 36-44, 48].

[11] [Wittgenstein, 1958], especially §§1-64.

As an alternative to these received resolutions of the liar, I now want to extend the liar metaparadox previously explained to provide semantic considerations that eliminate the paradox altogether in a new and distinctively deflationary way. The liar sentence L in the account I favor is itself a disguised contradiction. It therefore standardly validly supports the derivation of any other contradiction, as we see in the first horn of the liar paradox dilemma, necessitating the trivially valid deduction or default logical truth of the inference or counterpart material conditional in (i) that takes us from L to $S \wedge \neg S$ (and in particular from L to $\neg L$). This is an interesting result only in revealing the contradiction concealed within L, in somewhat the manner of the Moore pragmatic paradox sentence, 'It is raining, but I don't believe it'.[14]

The liar paradox in contrast can be seen to fail in the second dilemma horn. For, standardly, if the liar sentence L is judged to be a disguised contradiction that validly implies any other contradiction, then the negation of the liar sentence $\neg L$ must be a disguised tautology. I shall first try to justify these interpretations of L and $\neg L$, paying special attention to the intuitive rationale for regarding $\neg L$ as tautologous. Then I shall explain how this interpretation of the negation of the liar sentence in $\neg L$ blocks the standard valid inference from $\neg L$ to $S \wedge \neg S$ (and in particular from $\neg L$ to L) in the second liar paradox dilemma horn. By the same analogy, I conclude that the negation of the liar sentence is logically no more problematic and in many ways comparable to the pragmatic tautologous redundancy of the negation of the Moore sentence, 'Either it is not raining, or I believe that it is'.

---

[12] See [Priest, 1995, 167–171; 1987, 23–28].

[13] The dialethic logician need not blink at such a contradiction. Yet even [Priest, 1987, 104–105], writes: "...a necessary condition for entailment is truth-preservation from antecedent to consequent...Hence we may say that an entailment is false if it is possible for the antecedent to be true and the consequent false." Where a liar sentence is dialethically both true and false, it is not only possible but actually the case that the assumption (the liar sentence itself) of the liar dilemma inference is true (and false) while the conclusion (any contradiction) is both true and false. If by Priest's definition the liar paradox dilemma inference is deductively invalid (full stop), then the paradox is blocked. But if the inference works by virtue of matching the truth (not the falsehood) of the assumption with the truth (not the falsehood) of the conclusion, or by matching the (necessary) falsehood of the assumption and the truth or (and) falsehood of the conclusion, then the paradox goes through by deductively valid dilemma, though the true (and false) contradiction it entails may be semantically harmless. I can only conclude that I do not sufficiently understand the concept of deductive validity or entailment in a dialethic context to be able to assess its role and therefore to be able to endorse this type of paraconsistency solution to the liar paradox.

[14] The name 'Moore's Paradox' was given by Wittgenstein in response to a series of problems G.E. Moore mentioned in a paper delivered to the Moral Science Club at Cambridge University in 1944. Wittgenstein reportedly was more impressed by the paradox than Moore, who dismissed it as a psychological rather than logical absurdity. See [Monk, 1991, 544–547].

The negation of the liar sentence, 'It is false that this sentence is false', is standardly interpreted as logically equivalent to the affirmation, 'This sentence is true'. It is clear that the content of the liar sentence in denying its own truth is self-contradictory, for what the sentence says about itself implies that it is true if and only if it is false.[15] This analysis should entail that the negation of the contradictory liar sentence is a tautology. In what further intuitive sense, if any, is the negation of the liar sentence, 'This sentence is true', tautologous? The negation of the liar like any other sentence is (classically) either true or false. The negation of the liar sentence, on the other hand, cannot be false on pain of contradiction. For then it is logically equivalent to the liar sentence itself, 'It is false that this sentence is true' or 'This sentence is false', which by the above standardly implies an explicit logical contradiction. Hence, the negation of the liar sentence must be true. This is to say that the negation of the liar sentence is a disguised tautology in the same sense and ultimately for the same reason that the liar sentence is a disguised contradiction. That there can be such disguised contradictions and tautologies in natural language should not astonish anyone. To interpret the negation of the liar sentence in 'It is false that this sentence is false' or 'This sentence is true' as a tautology moreover engenders no further paradox, beyond surprising the uninitiated with the conclusion that a sentence that asserts its own truth in a kind of limiting case is necessarily trivially vacuously true.[16]

Although L as a disguised contradiction standardly validly implies $S \wedge \neg S$, $\neg L$ as a disguised tautology standardly implies only other tautologies, and consequently implies no contradictions. The liar paradox is stymied in this way at the second paradox dilemma horn. There is a confusion in the orthodox informal characterization of how the liar paradox is supposed to follow from the constructibility of the liar sentence. The semantic analysis of the content of the liar and its negation which has been sketched reveals the fallacy in the inference. The first part of the dilemma relying on inference or material conditional (i) above goes through well enough ((i) is valid or true). If the liar sentence is true, then undoubtedly it is false. If, on the contrary, the liar sentence is false, then it does not validly follow that the liar sentence itself is not true; rather, what follows logically is the literal negation of the liar sentence, which simply states, ‘This sentence is true’. *This* sentence, the *negation of the liar*, is a tautology, and therefore (necessarily) true. The assumption in the second liar paradox dilemma horn that the liar sentence is false, in applying the inference or material conditional in (ii), does not make the liar sentence itself true ((ii) is invalid or false). There is manifestly nothing paradoxical to be validly derived from the necessary truth of the disguised tautology in the negation of the liar sentence, ‘It is false that this sentence is false’ or ‘This sentence is true’. The liar dilemma is avoided by grasping the second horn, which produces only the potentially unexpected result that a sentence asserting its own truth is a tautology, and does not imply any logical antinomy by itself or in conjunction with the contradiction or necessary falsehood of the liar sentence.[17]

---

[15] The problem is not just that in the first liar paradox dilemma horn we obtain the conditional $L \rightarrow \neg L$. If we assume (that) L (is true) and from its content validly deduce or detach its negation $\neg L$ via inference or counterpart material conditional (i), then we derive the outright syntactical inconsistency $L \wedge \neg L$.

[16] Other examples abound: 'John has been a married bachelor for the last fifteen years'; 'Russell was the greatest philosopher of the 20th Century, and so was Wittgenstein'; 'First order logic with arithmetic is both syntactically consistent and deductively complete'.

I conclude that all the fuss made about the liar sentence and liar paradox is logically exaggerated. The liar sentence is not validly deducible in any otherwise sound application of language. If we want to avoid what we may perceive as a contradiction in the liar, despite the metaparadox conclusion that no such contradiction is validly entailed by the mere constructibility of the liar, then we should take pains not to utter it. Should philosophical semantics adopt a special policy to prevent us from lying? I do not think so; no more than semantics should legislate against the constructibility of just plain false sentences, or even more explicitly less controversially outright contradictory sentences such as $S \wedge \neg S$, or, say, the Moore sentence. There is a family of what might be called semantically degenerate constructions in languages, which include at one end of the spectrum explicit contradictions like $S \wedge \neg S$, with the liar and the Moore sentence situated somewhere closer to the other extreme. If we are interested in clarity of thought and language, and if we are interested in truth, then we should make an effort to avoid using these sentences. Similarly, if we are interested in making significant assertions, then we should avoid using disguised tautologies, like the negation of the liar sentence or pragmatically the negation of the Moore sentence. The mere fact that a given language is rich enough to permit the construction of these

---

[17] This method of resolving the liar paradox is not intended to provide a panacea for logical and semantic paradoxes generally. Compare Russell’s paradox in set theory, which validly deduces an explicit contradiction from two propositions that from a naive set theoretical viewpoint are apparently true by stipulation. The inference has this form:

$$(D)\quad \frac{\begin{array}{l} R = \{x \mid x \in x\} \\ R' = \{x \mid x \notin x\} \end{array}}{R' \in R \leftrightarrow R' \notin R}$$

The dissolution of the liar paradox by contrast is blocked because the concept of validity is itself standardly defined in terms of truth. The necessary falsehood of the liar paradox conclusion reflects back on the content of the liar sentence as the premise of the attempt to deduce the liar paradox from the mere constructibility of the liar. This is not typically a feature of logical or semantic paradox, as the Russell paradox shows. But it is a requirement of the liar paradox that in particular renders the effort to deduce the liar paradox from the liar sentence logically invalid in the second liar paradox dilemma horn, and hence logically and semantically innocuous.

oddities as limiting cases does not seem to constitute a good enough reason to prohibit them or to impose any special semantic restrictions on their formulation or interpretation.

The use of largely informal and dispensably quasi-formal logic in this analysis of the liar sentence shows the extent to which complicated formal symbolic notations and semantic relations can be avoided in the treatment of some of the most difficult problems in logic. The paradox disappears when we subject it to a thorough-going critical analysis using only the most modest and informal logical apparatus. To the extent that this is possible with other logical problems, the pragmatic principle I have proposed recommends that informal rather than formal logical methods be employed. That this is not always the most expedient use of logic is indicated by the second example I shall introduce. Here with respect to a notorious argument in philosophical theology it turns out to be more advantageous to use formal symbolic logic to discover a hidden logical fallacy in St. Anselm's ontological proof for the existence of God.

Anselm's ontological proof

Anselm's ontological proof for the existence of God has a precise modal structure. By formalizing the argument, it is possible to identify the intensional modal fallacy it contains. The deductive invalidity of Anselm's inference embodied in the fallacy defeats his argument, even if, contrary to Kant's famous objection, existence is included as a 'predicate' or identity-determining constitutive property of particulars.

Norman Malcolm in "Anselm's Ontological Arguments" distinguishes two forms of Anselm's inference.[18] Malcolm acknowledges that "There is no evidence that [Anselm] thought of himself as offering two different proofs."[19] Some commentators have understood what Malcolm refers to as the second ontological proof as Anselm's official or final formulation, interpreting the first version as a preliminary attempt to express or preparatory remarks for the demonstration's later restatement.[20] Anselm presents the so-called second ontological proof in *Proslogion* III:

> For there can be thought to exist something whose non-existence is inconceivable; and this thing is greater than anything whose non-existence is conceivable. Therefore, if that than which a greater cannot be thought could be thought to exist, then that than which a greater cannot be thought would not be that than which a greater cannot be thought — a contradiction. Hence, something than which a greater cannot be thought exists so truly that it cannot even be thought not to exist.

---

[18] [Malcolm, 1960, 41–62].

[19] [Malcolm, 1960, 45]. Malcolm explicitly states that Anselm does not separate the inferences, but maintains that the ontological argument can be better understood and defended if the two versions are distinguished.

[20] See [Schufreider, 1978, 40–45; Brecher, 1975, 140-146; Bencivenga, 1993, 113–123].

> And You are this being, O Lord our God. Therefore, Lord my God, You exist so truly that You cannot even be thought not to exist.[21]

This passage contains what I regard as the heart of Anselm's proof. For present purposes, I want to avoid the controversy of whether the text offers one or two distinct arguments. I shall therefore concentrate on this formulation of Anselm's argument, without trying to decide whether it is essentially the same as or relevantly different than related inferences about God's existence appearing elsewhere in Anselm's writings. I shall also follow Charles Hartshorne's recommendation in *The Logic of Perfection* by referring to the second statement of Anselm's ontological argument as a kind of modal proof, and of the 'irreducibly modal structure' of this form of Anselm's argument.[22] I nevertheless differ sharply from Hartshorne in interpreting the modality of Anselm's proof as intensional rather than alethic.

Where q abbreviates $(\exists x)Px$, that a perfect being or perfection exists, Hartshorne attributes this form to Anselm's proof:

Hartshorne's Formalization of Anselm's Alethic Modal Proof
for the Existence of God

1.	$q \rightarrow \Box q$	Anselm's principle
2.	$\Box q \vee \neg\Box q$	Excluded middle
3.	$\neg\Box q \rightarrow \Box\neg\Box q$	Becker's postulate
4.	$\Box q \vee \Box\neg\Box q$	(3 logical equivalence)
5.	$\Box\neg\Box q \rightarrow \Box\neg q$	(1 modal form of *modus tollens*)
6.	$\Box q \vee \Box\neg q$	(4,5 dilemma and detachment)
7.	$\neg\Box\neg q$	Perfection not logically impossible
8.	$\Box q$	(6,7 disjunctive dilemma)
9.	$\Box q \rightarrow q$	Modal axiom
10.	q	(8,9 detachment)

This is an elegant but defective derivation. Hartshorne presents Becker's Postulate in proposition (3) as though it were a universal modal truth. The principle in fact holds at most only in modal systems like S_5 with latitudinarian semantic transworld-accessibility relations.[23] There is furthermore a logical difficulty in a key assumption of the proof that renders the entire inference unsound.[24]

---

[21] [Anselm, 1974, I, 94]. [Anselm, 1945–51, I, 102–103], writes: "*Nam potest cogitari esse aliquid, quod non possit cogitari non esse; quod maius est quam quod non esse cogitari potest. Quare si id quo maius nequit cogitari, potest cogitari non esse: id ipsum quo maius cogitari nequit, non est id quo maius cogitari nequit; quod convenire non potest. Sic ergo vere est aliquid quo maius cogitari non potest, ut nec cogitari possit non esse. Et hoc es tu, domine deus noster. Sic ergo vere es, domine deus meus, ut nec cogitari possis non esse.*"

[22] [Hartshorne, 1962, 49–57]. I have converted Hartshorne's necessity operator 'M' to what is nowadays the more conventional symbol '$\Box$'. See [Hartshorne, 1965]. [Plantinga, 1974], offers another approach to the formal modal structure of Anselm's ontological argument.

[23] The principle Hartshorne identifies as Becker's Postulate, $\neg\Box q \rightarrow \Box\neg\Box q$, is logically equivalent to the characteristic axiom of modal S_5, more usually formulated as $\Diamond\alpha \rightarrow \Box\Diamond\alpha$. See [Purtill, 1966, 398].

[24] Hartshorne's proposition (1) is criticized by John Hick in [Hick and McGill, 1967, 349–352].

Proposition (5), which Hartshorne says follows from (1) as a modal form of *modus tollens*, is clearly false. Hartshorne glosses the assumption by maintaining that: "...the necessary falsity of the consequent [of (1)] implies that of the antecedent..."[25] The principle Hartshorne applies to proposition (1) to obtain (5) is thus: $(\alpha \rightarrow \beta) \rightarrow (\Box\neg\beta \rightarrow \Box\neg\alpha)$. This conditional is not generally true, as an obvious counterexample shows. Let $\alpha =$ Snow is red, and $\beta = 2 + 2 = 5$. Then the instantiation of $\alpha \rightarrow \beta$ in Snow is red $\rightarrow 2 + 2 = 5$ is true by default, if the $\rightarrow$ conditional is interpreted classically, since it is false that snow is red. It is false that $\Box\neg\beta \rightarrow \Box\neg\alpha$ in the instantiation $\Box\neg(2 + 2 = 5) \rightarrow \Box\neg$(Snow is red), however, because although it is true that $\Box\neg(2 + 2 = 5)$, it is false that $\Box\neg$(Snow is red). If Hartshorne tries to avoid the counterexample by interpreting the conditional $\rightarrow$ nonstandardly in a relevance logic, then the proof is either deprived of excluded middle in proposition (2), or logically disabled in its crucial inference from the disjunction in (2) to the conditional in (3).[26]

The problem is magnified in Hartshorne's application of this 'modal form of *modus tollens*' to proposition q in proof step (5). Consider that if it is necessary that it is not necessary that snow is white, it by no means follows that it is necessary (particularly because *a fortiori* it is not actually the case) that snow is not white. This refutes the general truth of (5) along with the general principle on which it is supposed to depend. What about proposition (5), in Hartshorne's specific interpretation, according to which q means $(\exists x)Px$, that a perfect being or perfection exists? After all, proposition (1) is also not generally true, but at most, Hartshorne only believes, when q abbreviates $(\exists x)Px$. Here is a dilemma. By Hartshorne's appeal to excluded middle, either proposition q or its negation is true, $q \vee \neg q$. If $\neg q$, then Anselm's ontological argument is logically unsound on any interpretation. If q, then, given Hartshorne's other assumptions in his formalization, the following inference holds:

Classical Logical Triviality of Hartshorne's Modal Principle (5)

1.	q	Assumption
2.	$q \rightarrow \Box q$	Anselm's principle (Hartshorne)
3.	$\Box q \rightarrow \Diamond\Box q$	Modal axiom ($\alpha \rightarrow \Diamond\alpha$)
4.	$\Diamond\Box q \rightarrow \neg\Box\neg\Box q$	Modal duality ($\Diamond\alpha \longleftrightarrow \neg\Box\neg\alpha$)
5.	$q \rightarrow \neg\Box\neg\Box q$	(1-4 hypothetical syllogism)
6.	$\neg\Box\neg\Box q$	(1,5 detachment)

If q is true, as Hartshorne's conclusion (10) states, that a perfect being or that perfection exists, then Hartshorne's proposition (5) is logically trivial. For then

---

See also [Brecher, 1976, 140–146; Plantinga, 1961, 93–101. Hartshorne, 1967, 290–309], replies to Purtill on the proper modal logical concept of necessity as it pertains to Anselm's efforts to prove the existence of God. Another kind of informal criticism is offered by [Nelson, 1963, 235–242].

[25] [Hartshorne, 1962, 51].

[26] On the unavailability of excluded middle and disjunctive syllogism in relevance logic, see [Read, 1988, 60].

the antecedent of (5) is false, making $\Box\neg\Box q \rightarrow \Box\neg q$ an empty truism on the classical interpretation of the conditional $\rightarrow$. This makes it equally (classically) uninterestingly true both that $\Box\neg\Box q \rightarrow \neg\Box\neg q$ and $\Box\neg\Box q \rightarrow \Box q$. Since, however, Hartshorne's formalization relies on proposition (5) in its derivation of (6) from (4) and (3), it follows that the conclusion of Anselm's proof that God exists is true only if Anselm's proof for the conclusion as Hartshorne interprets it is (classically) logically trivial.

If Hartshorne has correctly represented Anselm's modal ontological argument, then the proof is clearly in bad shape. Hartshorne's reconstruction makes no use and takes no notice of Anselm's idea about the conceivability of God as a being that than which none greater can be conceived, nor does Hartshorne construe Anselm's proof as the *reductio ad absurdum* Anselm intends when he argues as above: '...if that than which a greater cannot be thought could be thought to exist, then that than which a greater cannot be thought would not be that than which a greater cannot be thought — a contradiction'.[27] The principle of charity therefore requires an effort to identify an alternative formalization of Anselm's modal argument that avoids Hartshorne's commitment to the manifestly unsound principle in proposition (5).

I now want to apply some of the symbolizations Priest recommends in formalizing the argument from Anselm's discussion of God's inconceivability to Anselm's ontological proof for God's existence. Like Priest, I use an indifferently definite or indefinite description operator δ, and I adopt an informal ontically neutral interpretation of the quantifiers, expressing existence by means of a predicate, $E!$ (E-shriek). I also follow Priest in symbolizing Anselm's relation of 'being greater than' by the convenient predicate '$>$'. However, I revise Priest's conceivability operator to extend its scope. I allow the τ-operator to range over a greater-than relation to express the conceivability that the relation holds between two objects, instead of merely attaching to an individual object term as a way of expressing the corresponding object's conceivability. Then I can logically represent Anselm's definition of God as that than which none greater is conceivable by the expression $\tau(y > x)$ in $g = \delta x \neg\exists y \tau(y > x)$.[28]

In contrast with Hartshorne, I will not emphasize the proof's alethic modality in the proposition that if a perfect being exists then it necessarily exists. Instead, I wish to call attention to the proof's hitherto neglected intensional modality implied by Anselm's reliance on the concept of God as a being than which none greater is *conceivable*. This formalization more accurately reflects Anselm's thinking in the ontological proof, which I take to be an improvement over Hartshorne's. I argue that the intensional modality implied by wide-scope conceivability in Anselm's definition of God renders the ontological argument deductively invalid. For as such it requires a violation *salva non veritate* of the extensionality of Priest's definite or indefinite description operator. Then, even if Kant's objection that existence is not a 'predicate' or identity-determining constitutive property is overturned, the logical structure of Anselm's argument on the most charitable reconstruction nevertheless fails by virtue of instantiating an intensional modal fallacy.

---

[27] See [Henry, 1967, 143].

[28] The difference is not trivial, but concerns whether Anselm is to be understood as saying merely that God is that than which there is nothing conceivable that is greater, or that than which it is not conceivable that there be anything greater. Compare Priest's definition in [1995, 62, n. 2]: "Let τx be 'x is conceived'. Then God (g) may be defined as $\delta x \neg\exists y(\tau y \wedge y > x)$. (Quantifiers, note, are not existentially loaded.) Let $\varphi(x)$ be the second-order condition $Ex \wedge \forall P(P \neq E \rightarrow (Px \leftrightarrow Pg))$, where E is existence. Then the claim is $\forall x(\neg Eg \wedge \varphi(x) \rightarrow x < g)$." See [Priest, 1979, 5–21]. I make a similar plea for the ontic neutrality of the quantifier in noncircular reconstructions of Anselm's proof in [Jacquette, 1994, 231–240].

The ontological proof is formalized by the following inference, in which two new principles are adduced. The first thesis maintains the extensionality of δ, stating that anything identical to a definitely or indefinitely described object has whatever properties are attributed to the object by the description. The second is a conceivable greatness thesis, which states that there is always an existent or nonexistent object that is conceivably (if not also in fact actually) greater in Anselm's sense than any nonexistent object. Anselm's argument can now be symbolized in this way:

Anselm's Intensional Modal Proof
for the Existence of God

1.	$g = \delta x \neg \exists y \tau (y > x)$	Definition g
2.	$\forall x(\delta y \ldots y \ldots = x \rightarrow \ldots x \ldots)$	Extensionality δ
3.	$\forall x(\neg E!x \rightarrow \exists y \tau (y > x))$	Conceivable greatness
4.	$\neg E!g$	Hypothesis for *reductio*
5.	$\neg E!g \rightarrow \exists y \tau (y > g)$	(3 instantiation)
6.	$\exists y \tau (y > g)$	(4,5 detachment)
7.	$\neg \exists y \tau (y > g)$	(1,2 instantiation)
8.	$E!g$	(4,6,7 *reductio ad absurdum*)

As in Hartshorne's formalization, Anselm is interpreted as drawing inferences from propositons containing modal contexts. Here, as opposed to Hartshorne's rendition, the modal contexts in question are intensional rather than alethic, expressing the intentionality of wide-scope conceivability in the interpolated conceivable greatness principle.

The proof has several advantages over previous attempts to symbolize Anselm's ontological argument. The argument as reconstructed is extremely compact, reflecting approximately the same level of complexity as Anselm's original prose statement. Anselm's definition of God as that than which none greater is conceivable is explicit in proposition (1). The *punctum saliens* of Anselm's proof, that if God does not exist, then there is after all something conceivably greater than God, is featured prominently in proposition (5), derived via instantiation from the more general conceivable greatness thesis in (3). Finally, unlike Hartshorne's version, the proposed formalization explicitly represents Anselm's proof as a *reductio.* To define God as that than which none greater is conceivable, and to suppose that God does not exist, is to be embroiled in logical contradiction, if, as Anselm seems to assume, we can always conceive of something greater than anything that does not actually exist.

The proposed method of formalizing Anselm's argument makes it easy to discover the proof's logical weakness. The problem arises in proposition (7). The conceivability context, $\tau\langle\underline{\quad}\rangle$, is modal, because it represents an intensional mode of whatever sentence or proposition is inserted. The context's intensionality is seen in the fact that coreferential singular denoting terms and logically equivalent propositions cannot be freely intersubstituted in the context *salva veritate*. Thus, whereas virtually any short and sweet tautology is conceivable, not every infinitely long or monstrously complicated tautology logically equivalent to it is also thereby conceivable.

The intensionality of conceivability invalidates the inference from (1) and (2) to (7), by requiring the substitution of g in definition (1) for the definitely or indefinitely description-bound variable x by the extensionality of δ in principle (2). The Kantian objection to Anselm's proof can also be pinpointed in this symbolization by linking it to the conceivable greatness thesis in proposition (3). Kant refutes the ontological argument in the section on 'The Ideal of Pure Reason' in the *Critique of Pure Reason* A599/B627-A600/B628. Kant's claim that existence has no part in the identity-determining constitutive properties of 100 real or unreal gold Thalers challenges the principle in (3) that we can always conceive of something greater than any nonexistent object. Anselm's ontological argument cannot succeed if relative greatness is judged only by a comparison of the constitutive properties that make two or more objects the particular objects they are, to the exclusion of all extraconstitutive properties that categorize such objects' ontic status as existent or nonexistent.

Kant is undoubtedly right to draw this conditional conclusion. The important question is whether Kant is entitled to claim that existence is not a 'predicate' or identity-determining constitutive property. There are intriguing proposals for avoiding Kant's 100 gold Thalers criticism in the philosophical literature, and for reconciling ourselves to accepting existence as a 'predicate' in the one special logically unique case of God.[29] Without entering into the merits of these replies to Kant, it is worth remarking that even if Kant's objection to the assumption in (3) is forestalled, the intensional modal structure of Anselm's argument remains deductively invalid.[30] The intensional fallacy arises in Anselm's proof because of its attempt to apply the description context extensionality principle in assumption (2) at inference step (7) to the definition of God in assumption (1). Since Kant's conceptual-metaphysical or grammatical refutation of the ontological argument is controversial, the deductive logical invalidity entailed by the argument's committing the intensional fallacy may now appear to be the more fundamental and decisive objection to Anselm's ontological proof. Anselm might avoid venially transgressing Kant's injunction against treating existence as a 'predicate'. The intensional fallacy in Anselm's modal ontological argument for the existence of God is nevertheless its more deadly cardinal sin. The logical error in Anselm's proof has been overlooked in the history of purely informal discussions of the inference. It is only by formalizing the logical structure of Anselm's argument that it becomes possible to identity its deductive invalidity.

---

[29] See [Shaffer, 1962, 307–325; Engel, 1963, 20–35; Plantinga, 1966, 537–546].

[30] An inference in this same style of logical notation that is subject to Kant's objection can also be formalized. Instead of relying on the extensionality of δ, the method appeals to what Priest calls the Characterization Principle. The argument takes the following form:

4 CONCLUSION

The examples I have considered are not intended as the only applications of the pragmatic principle I have endorsed for a partnership between informal and symbolic logic. They are meant to be illustrative only in an area where many other solutions or attitudes are possible. The general point is partly to voice a plea for tolerance in logical investigations, to recognize and take full advantage of the continuum of methods available in logic, any of which might be useful and none of which should be overlooked because of ideological or methodological prejudice. In criticizing intelligent but faulty arguments, it is often necessary to combine complementary informal with symbolic logical techniques in cooperative analytic partnerships.[31]

1.	$\varphi\delta x(\varphi x))$	Characterization
2.	$g = \delta x \neg\exists y\tau(y > x)$	Definition g
3.	$\forall x(\neg E!x \rightarrow \exists y\tau(y > x))$	Conceivable greatness
4.	$\neg E!g$	Hypothesis for *reductio*
5.	$\neg E!g \rightarrow \exists y\tau(y > g)$	(3 instantiation)
6.	$\exists y\tau(y > g)$	(4,5 detachment)
7.	$\lambda z[z = \neg\exists y\tau(y > z)]\delta x\neg\exists y\tau(y > x)$	(1 instantiation)
8.	$\neg\exists y\tau(y > \delta x\neg\exists y\tau(y > x))$	(7 λ-equivalence)
9.	$\neg\exists y\tau(y > g)$	(2,8 Substitution of identicals)
10.	$E!g$	(4,6,9 *reductio ad absurdum*)

Kant's 100 gold Thalers criticism balks at the attempt in step (3) to give an instantiation of the Characterization Principle. The problem is that nonexistence like existence for Kant is not a 'predicate', which is to say that nonexistence is not a constitutive identity-determining property. There is moreover a sense in which this alternative method of formalizing Anselm's proof falls back into the same problem as symbolizations involving the extensionality of descriptor δ. By itself, $\neg\exists y\tau(y > x)$ with its unbound variable is not well-formed, and does not designate a property or even a relational property term. To obtain a substitution instance of Characterization as in (8) above, it is necessary to resort as in (7) to instantiation via λ-conversion. Yet there is a logical equivalence between λ-abstractions and descriptions, as expressed in this untyped statement intuitively equating a formal abstraction with the or a definitely or indefinitely described property satisfying certain conditions: $(\forall x)(\lambda y[\ldots y \ldots]x \leftrightarrow \delta z(\forall y)(zy \leftrightarrow (\ldots y \ldots)x))$. Applying λ-equivalence to proposition (7) to deduce step (8) thereby again presupposes the extensionality of δ. I was led to consider this version of Anselm's proof in responding to questions raised by Priest in personal correspondence.

[31] A version of one main part of this essay was presented at the Conference on Truth, sponsored by the Inter-University Centre of Dubrovnik, Croatia, held in Bled, Slovenia, June 3–8, 1996, see [Jacquette, 1997]. I am grateful to John Biro and Brian McLaughlin for useful comments and criticisms. The essay appeared subsequently as [Jacquette, 2000]. The material on Anselm is extracted with revisions from my essay on "Conceivability, Intensionality, and the Logic of Anselm's Modal Argument for the Existence of God", *International Journal for Philosophy of*

ACKNOWLEDGEMENTS

I am grateful to the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS), Royal Netherlands Academy of Arts and Sciences (KNAW), for supporting this and related research in philosophical logic and philosophy of mathematics during my Resident Research Fellowship at the Institute in 2005-2006.

BIBLIOGRAPHY

[Anselm, 1945–51] Anselm of Canterbury. *Opera omnia, ad fidem codicum recensuit Franciscus Salesius Schmitt.* Edinburgh: Nelson & Sons, 1945-51, 6 vols.), I. 1945-51.

[Anselm, 1974] Anselm of Canterbury. *Anselm of Canterbury.* Trans. Jasper Hopkins and Herbert Richardson. Toronto and New York: Edwin Mellen Press, 4 vols., I. 1974.

[Barwise and Etchemendy, 1987] . Barwise and J. Etchemendy. *The Liar: An Essay on Truth and Circularity.* Oxford: Oxford University Press. 1987.

[Bencivenga, 1993] E. Bencivenga. *Logic and Other Nonsense: The Case of Anselm and his God.* Princeton: Princeton University Press. 1993.

[Brecher, 1975] R. Brecher. Hartshorne's modal argument for the existence of God. *Ratio*, 17, 140-146. 1975.

[Engel, 1963] S. Morris Engel. Kant's 'refutation' of the ontological argument, *Philosophy and Phenomenological Research*, 24, 20-35. 1963.

[Hartshorne, 1962] C. Hartshorne. *The Logic of Perfection and Other Essays in Neoclassical Metaphysics.* LaSalle: Open Court Publishing Co. 1962.

[Hartshorne, 1965] C. Hartshorne. *Anselm's Discovery: A Re-Examination of the Ontological Proof for God's Existence.* LaSalle: Open Court Publishing Co. 1965.

[Hartshorne, 1967] C. Hartshorne. Necessity, *Review of Metaphysics*, 21, 290-309. 1967.

[Henry, 1967] D. P. Henry. *The Logic of Saint Anselm.* Oxford: The Clarendon Press. 1967.

[Hick and McGill, 1967] J. Hick and A. C. McGill, ed. *The Many-Faced Argument: Recent Studies on the Ontological Argument for the Existence of God.* New York: The Macmillan Company. 1967.

[Jacquette, 1994] D. Jacquette. Meinongian logic and anselm's ontological proof for the existence of God, *The Philosophical Forum*, 25, 231-240. 1994.

[Jacquette, 1997] D. Jacquette. Conceivability, intensionality, and the logic of anselm's modal argument for the existence of God, *International Journal for Philosophy of Religion*, 42, 163-173. 1997.

[Jacquette, 2000] D. Jacquette. Liar paradox and metaparadox, *Sats: Nordisk Tidsskrift for Filosofi*, 1, 93–104. 2000.

[Kant, 1965] I. Kant. *Critique of Pure Reason.* Trans. Norman Kemp Smith. New York: St. Martin's Press. 1965.

[Kripke, 1975] S. A. Kripke. Outline of a theory of truth, *The Journal of Philosophy*, 72, 690-716. 1975.

[Malcolm, 1960] N. Malcolm. Anselm's ontological arguments, *The Philosophical Review*, 69, 41-62. 1960.

[Martin, 1970] R. L. Martin, ed. *The Paradox of the Liar.* New Haven: Yale University Press. 1970.

[Martin, 1984] R. L. Martin, ed. *Recent Essays on Truth and the Liar Paradox.* Oxford: The Clarendon Press. 1984.

[Monk, 1991] R. Monk. *Wittgenstein: The Duty of Genius.* New York: Penguin Books. 1991.

[Nelson, 1963] J.O. Nelson. Modal logic and the ontological proof of God's existence, *Review of Metaphysics*, 17, 235-242. 1963.

[Plantinga, 1961] A. Plantinga. A valid ontological argument?, *The Philosophical Review*, 70, 93-101. 1961.

Religion, 42, 1997, pp. 163-173.

[Plantinga, 1966] A. Plantinga. Kant's objection to the ontological argument, *The Journal of Philosophy*, 63, 537-546. 1966.

[Plantinga, 1974] A. Plantinga. *The Nature of Necessity.* Oxford: The Clarendon Press. 1974.

[Priest, 1979] G. Priest. Indefinite descriptions, *Logique et Analyse*, 22, 5-21. 1979.

[Priest, 1987] G. Priest. *In Contradiction: A Study of the Transconsistent.* Dordrecht and Boston: Martinus Nijhoff Publishers. 1987.

[Priest and Routley, 1989] G. Priest and R. Routley. First historical introduction: a preliminary history of paraconsistent and dialethic approaches, 12-13, 22-23, 36-44, 48. In Priest, Routley and Jean Norman, Ed. *Paraconsistent Logic: Essays on the Inconsistent.* Munich and Vienna: Philosophia Verlag. 1989.

[Priest, 1995] G. Priest. *Beyond the Limits of Thought.* Cambridge: Cambridge University Press. 1995.

[Purtill, 1966] R.L. Purtill. Hartshorne's modal proof, *The Journal of Philosophy*, 63, 397-408. 1966.

[Read, 1988] Stephen Read. *Relevant Logic: A Philosophical Examination of Inference.* Oxford: Basil Blackwell. 1988.

[Rüstow, 1987] A. Rüstow. *Der Lügner: Theorie, Geschichte und Auflösung.* New York: Garland Publishing, Inc. 1987.

[Schufreider, 1978] G. Schufreider. *An Introduction to Anselm's Argument.* Philadelphia: Temple University Press. 1978.

[Shaffer, 1962] J. Shaffer. Existence, predication and the ontological argument, *Mind*, 71, 307-325. 1962.

[Skyrms, 1970] B. Skyrms. Return of the liar: three-valued logic and the concept of truth, *American Philosophical Quarterly,* 7, 153-161. 1970.

[Snow, 1964] C.P. Snow. *The Two Cultures: And a Second Look.* New York: New American Library. 1964.

[Tarski, 1956] A. Tarski. The concept of truth in formalized languages, 152-278. In Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938.* Trans. J.H. Woodger. Oxford: The Clarendon Press. 1956.

[Wittgenstein, 1922] L. Wittgenstein. *Tractatus Logico-Philosophicus.* Ed. C.K. Ogden. London: Routledge & Kegan Paul. 1922.

[Wittgenstein, 1958] L. Wittgenstein. *Philosophical Investigations*, 3rd ed. Ed. G.E.M. Anscombe. New York: Macmillan Publishing Co., Inc. 1958.

VAGUENESS AND THE LOGIC OF ORDINARY LANGUAGE

Roy A. Sorensen

Suppose that "some mathematical creature from the moon" surveys the human body. The alien moves from the left foot to the right foot, from the left hand to the right hand, from the left lung to the right lung. Such symmetry! Upon reaching the heart, the alien infers that there is also a right heart. G. K. Chesterton draws a moral:

> The real trouble with this world of ours is not that it is an unreasonable world, nor even that it is a reasonable one. The commonest kind of trouble is that it is nearly reasonable, but not quite. Life is not illogicality; yet it is a trap for logicians. It looks just a little more mathematical and regular than it is; its exactitude is obvious, but its inexactitude is hidden; its wildness lies in wait. [1927, 148]

The sorites paradox seems to be one of the traps about which Chesterton warns. When a logician looks at a mountain, he reasons:

Base step: It is safe to ascend one meter.

Induction step: If an ascent of n meters is safe, then so is an ascent of $n + 1$ meters.

Conclusion: It is safe to ascend all the way up the mountain.

As the light fades and the wind begins to howl, the logician finds himself stranded.

Where did he go wrong?

1 THE EPISTEMIC DIAGNOSIS

Socrates said we should follow the argument wherever it leads. But the mathematician's first rule for checking his calculation is to see whether the conclusion is plausible. If it is not, he reviews his logic. The above argument is an instance of the valid argument form of mathematical induction. Only the truth of the premises remains to be checked. The base step is obviously true. Therefore, the induction step must be false. Epistemicists [Sorensen, 2001; Williamson, 1994] conclude that there must have been a last safe meter.

Which was it? There is no way to know. The near perfect symmetry of 'An ascent of n meters is safe' and 'An ascent of $n+1$ meters is safe' make it impossible to affirm one while denying the other.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

Most checking procedures only tell you that there is an error somewhere. Often we can go on to specify the mistake. The epistemicist admits that he has no prospect of singling out the last safe step of the mountain climbing sorites. Happily, such specificity is not needed to solve the sorites paradox. The epistemicist has disproved the induction step by means that satisfy mathematical standards of rigor. What more could you want?

"Sanity!" answer most philosophers. They reject epistemicism as a fanatical overextension of classical logic. In the following pages I will review their reasons for restricting or revising or rescinding classical logic. Most of these challenges either spring from insights about vagueness or prompt such insights. Those who wish to learn how logic relates to language can do little better than to explore such confrontations between theory and practice. However, my overarching thesis is negative: vagueness provides no grounds to restrain classical logic. I believe that our best prospect for understanding ordinary language is through a resolute application of classical logic. I leave it to the reader to judge whether the epistemicist crosses the unclear line between persistence and dogmatism.

2 DOES LOGIC APPLY TO THE ORDINARY WORLD?

In "Vagueness" Bertrand Russell [1923] declared that logic applies to the Platonic heavens, not to the terrestrial world. Plato regards the ordinary world as a shadow of the world of the Forms. This picture accords with the fact that 'vague' is a pejorative term while 'precise' is honorific.

Fredrich Nietzsche agreed that logic is only suited to the forms but he also regarded the Platonic realm as a rationalist fantasy:

> Logic, too, also rests on assumptions that do not correspond to anything in the real world, e.g., on the assumption that there are equal things, that the same thing is identical at different points in time: but this science arose as a result of the opposite belief (that such things actually exist in the real world). And it is the same with mathematics, which would certainly never have arisen if it had been understood from the beginning that there is no such thing in nature as a perfectly straight line, a true circle, and absolute measure. [1878, section 11].

The Romantics who preceded Nietzsche believed that the logician projects categories and boundaries on nature.

The projection of specificity is deeply rooted and not restricted to intellectuals. The neuroscientific term 'pareidolia' denotes the brain's penchant for interpreting vague images as specific images. In 1976 the Viking Orbiter radioed back many images from Mars. Among those from the Cydonia region, one depicted a rock formation that strongly resembled a human face staring straight up into the heavens. The public interpreted this as evidence of intelligent life on Mars. NASA was pressured to take more photographs on the next mission.

The Romantics believed that mountains are vague objects. Why is there no way to tell where the mountain ends and the countryside begins? Because there is nothing to tell! Ordinary language is vague because reality is vague. Since Mount Everest lacks a boundary, so does the name 'Mount Everest'.

If anything, ordinary language is too precise. It fabricates boundaries so that we can have terms for our syllogisms. The colors of the rainbow are continuous but ordinary language imposes a conceptual scheme of discrete color terms. We mistake this artificial grid as a natural feature of the landscape.

The Romantics' favorite musical instrument was the Aeolian harp (the "wind harp") because it is played without human intervention. The audience hears nature directly. The Romantics believed that a receptive poet lets nature's message stream through his lips.

And what do the Romantic poets tell us? That reality is continuous rather than discrete, that it is holistic rather than piecemeal, that things blur into each other. There is enough regularity to hypnotize generalizers. But the line, the square, and the cube cannot domesticate the coast, the sea, and the sky.

3 VAGUENESS IN THE PLATONIC HEAVENS

Russell, Nietzsche, and the Romantics overestimate the tameness of abstractions. Mathematics teems with perverse exceptions to elegant rules. It has to be a reservoir of broken patterns to model rational disappointments. Johannes Kepler initially believed that the orbit of Mars is circular because that was the simplest model. When the data did not quite fit, Kepler recovered by acquiescing to the added complexity of an ellipse. Kepler may have first felt that Mars was being obstreperous. But that was just a projection of his frustration.

I once cursed the inexactitude of an old set of wrenches because none of them would snugly fit the bolts of my new bicycle. My anger was ill founded: the old wrenches were in the imperial system while the new bicycle bolts were in the metric system. 'Exact' is relative. When Chesterton speaks of the subtle inexactitude of nature, he mischaracterizes nature as *intrinsically* inexact.

As the level of specificity increases, there are more potential counterexamples. We should expect simple models to choke on fine-grained input. Even Kepler's hypothesis of an elliptical orbit must be revised for the purpose of planetary exploration. After all, the sun and Mars are not point-masses (as Newton understood when he derived Kepler's laws of planetary motion from the inverse square law of gravitational attraction).

The wildness of *nature* cannot be the sole source of sorites arguments. For sorites arguments apply to abstract entities. Take Wang's paradox: 0 is a small number. If n is a small number, then $n+1$ is a small number. Therefore, a billion is a small number.

The culprit is vagueness rather than concreteness. Every sorites argument has a vague inductive predicate. This vagueness does not cause any trouble when the induction sticks to clear instances of the vague predicate. After all, the following argument is sound: A billion is a large number. If n is a large number, then $n + 1$ is a large number. Therefore, a trillion is large number.

Mathematicians do not try to separate good but vague mathematical inductions from sorites arguments. They ignore the whole lot by invoking their traditional ban on vagueness in proofs. The ban began in Greek mathematics and contributes to the decisiveness of the Western mathematical tradition. The poetical mathematicians of ancient India have flashes of brilliance but do not consolidate their gains.

Statements that appear in proofs do not exhaust the set of truths about mathematical entities. Teachers who study arithmetical reasoning in children apply many psychological principles about numbers: *Small numbers are easier to add than large numbers, Hindu-Arabic numerals let us reason about large numbers more efficiently than the Roman numerals,* and so on. These relational statements are as much about numbers as they are about children.

Since contemporary mathematicians expurgate vague predicates from proofs, they purchase precision at the price of completeness. Plato did not have this option because his theory of Forms was presented as a *comprehensive* answer to 'How do we use general terms?' Consequently, vague predicates apply to Plato's abstract entities. In his middle period, at least, (before worries about the third man argument in the *Parmenides*) Plato maintains that the Forms are self-predicational: Goodness is itself good. Largeness is itself large. Plato also uses vague vocabulary to describe how the Forms relate to particulars and how Forms relate to one another. For instance, he arranges Forms on a scale from lower to higher Forms.

Chesterton assumes that near misses of orderliness are always contingent. The human body could have been perfectly symmetrical. The earth could have been perfectly round instead of bulging at the equator. The solar cycle could have been an integral multiple of the lunar cycle.

But some near misses are necessary truths. The Pythagoreans believed that numbers governed the universe. So they assumed that the sides of a triangle were always commensurable. They were frustrated by the isosceles right triangle. They could find ratios that ever more closely approximate the relationship the sides bear to the hypotenuse. But the sequence of nearer and nearer misses never hit the target.

In 1974 advances in aperiodic tiling refreshed the theme of subtle inexactitude. If an infinite surface is covered with Roger Penrose's "kites" and "darts", every finite region is contained somewhere inside every other pattern. But the pattern never quite repeats. Martin Gardner [1989, 9] cites the continuation of the Chesterton quotation: "Everywhere there is a silent swerving from accuracy by an inch that is the uncanny element in everything . . .a sort of secret treason in the universe." [1927, 149] But the aptness of the quotation is imperfect because the "treason" committed by Penrose's tiles is a necessary truth. The swerve is predictable by *a priori* reasoning. It is not an *a posteriori* stab in the back.

Philosophers who posit two worlds exaggerate the differences between residents of those worlds. They neglect entities that bridge the worlds. For instance, some abstract entities, such as the set that contains just Mt. Everest, depend on concrete individuals for their existence. Since this set exists contingently, not all abstract objects are necessary beings. If there were no Mt. Everest, there would not be the singleton set containing Mt. Everest. If Mt. Everest were a vague object, then {Mt. Everest} would be a vague object. Whatever vagaries affect Mt. Everest echo with {Mt. Everest}, and {{Mt. Everest}} and so on.

4 DOES LOGIC APPLY TO ORDINARY LANGUAGE?

There is no difference in how well logic applies to abstract objects and concrete objects, However, a determined critic of classical logic could still conclude that logic applies to *neither*.

There is a way of interpreting this radical position so that it comes out technically correct: Logic applies to propositions rather than things. Logic is a theory of entailment and that is a relationship between propositions.

Once we make this semantic ascent, anti-epistemicists can reframe the initial asymmetry in linguistic terms: Logic applies to the precise language of mathematicians, not the vague language of ordinary speakers.

However, it is self-stultifying to *deduce* that logic does not apply to any reasoning. Meta-reasoning is reasoning. Inapplicability arguments use vague predicates such as 'inapplicable' and 'vague'.

To avoid self-defeat, we must assume that logic does apply ordinary reasoning. The new question is this: How do the perfectly round wheels of logic get traction in the mud of ordinary language?

5 VAGUENESS AS A PRE-LOGICAL PHENOMENON

Susan Haack's [1974, ch. 6] answer is that classical logic only applies to vague predicates *after* they have been precisified. This preparation is standard operating procedure in the case of ambiguity.

If vagueness is hyper-ambiguity, then we can directly apply the maxim 'Disambiguation precedes logic'. However, vagueness differs from ambiguity in key respects [Sorensen, 1998]. Concepts can be vague but cannot be ambiguous (since words are disambiguated by concepts). There could be vagueness in the language of thought even though there could not be any ambiguity. One cannot wittingly believe a sentence with an unresolved ambiguity. However, one can wittingly believe a statement with unresolved vagueness. Ambiguity confers discretion over what one means. Vagueness does not. Lastly, the identification of vagueness with ambiguity is undermined by the fact that there is higher order vagueness but no higher order ambiguity. (Ambiguity about ambiguity collapses into ambiguity.) 'Vague' is vague but 'ambiguous' is not ambiguous, so vagueness cannot be ambiguity.

The amount of vagueness in a natural language exactly equals the amount of vagueness of any other natural language. This follows from the inter-translatability of languages. You cannot resolve a borderline case by translation into a different language. The constancy of vagueness also follows from the psychological premise that there is a single language of thought, "mentalese", underlying all natural languages.

The amount of ambiguity does vary across languages and within stages of a language. Consequently, translation can remove an ambiguity. Surprisingly, the speech of children becomes more ambiguous as their speech matures into adult form. For they master irregularities of omission. After a boy learns the *–ed* rule for past tense, he overregularizes verbs. He says 'On Thursday I putted the trash out'. After he stops overregularizing, he says 'On Thursday I put the trash out'. But notice that this "improvement" is ambiguous. It could mean that last Thursday he took the trash out or next Thursday he will take the trash out or mean that he habitually takes the trash out every Thursday. English has about twenty-five irregular verbs (*put, cut, set*) that don't change their forms in the past tense. When children become sufficiently exposed to the irregular usage, their used of the *–ed* rule is blocked and their speech becomes more ambiguous.

Granted, disambiguation and precisification are salient features of scientific language. Scientists replace qualitative terms (hot) with comparative terms (hotter than) and comparative terms with degree terms (n degrees centigrade). Rudolph Carnap [1950] looked forward to the day when language was so precise that the sorites paradox could no longer be expressed.

Are precisifications cumulative? Often they do promote peace and stability. Legislators prevent disputes over borderline cases by imposing sharp cut-offs for adulthood, death, and citizenship. Law accumulates definitions because they promote predictability.

However, the stabilizing effect is not universal. Discord is sharpened when indefinite borderline cases are precisified into definite borderline cases. Hundreds of islets in the Aegean Sea were not mentioned by name in the Treaty of Lausanne (or by later treaties that ceded the eastern islands to Greece). The Greeks regard the islets as dependent on the major islands and so part of their territory. The Turks only recognize Greek sovereignty over the islands specifically named in the Treaty. The Turks were vague about which islets were part of the named islands and which islets were in "the grey zone'. This changed in 1998 when the Turkish prime minister specified that Agathonisi, Farmakonisi, Fourni, and Pserimos were borderline cases. He asked the European Union not to fund environmental projects until the sovereignty of those islets was settled.

One may complain that the Aegean example is unfair because the precisification was incomplete. Further precisification, perhaps by the negotiation proposed by the Turks, would bring about the peace promised by precisification.

So let me mention a case that does not turn on the unearthing of underlying borderline cases. In 1910, residents of Taff's Well were outraged when cartographers classified Garth Mountain as a hill. The villagers had taken pride in it being the first mountain inside of Wales — especially since the English have no mountains. The two representatives of His Majesty's Ordnance Survey explained that a mountain must have an elevation of 1000 feet. They measured Garth Mountain at 986 feet. The locals detained the cartographers and built a twenty-foot mound at the summit. The incident became the basis for the 1995 movie "The Englishman Who Went Up a Hill But Came Down a Mountain".

The type of precisification practiced by scientists and administrators is vagueness reduction, not the total elimination of vagueness. Consider all the residual vague terms in the definition "A mountain is an area of land that stands up at least 1000 feet from the surrounding countryside." If cartographers tried to excise *all* the vague terms, the concept would be disfigured beyond recognition. However, only complete precisification makes a term invulnerable to the sorites.

Scientists and administrators are not trying to solve the sorites paradox so the residual vagueness does not frustrate their goals. The incompleteness is a problem for philosophers who want to solve the sorites paradox with these precisifications.

Carnap also underestimates our success at reasoning with vague terms. The original sense of 'mountain' co-exists with its precisified offspring. If we were not able to reason with 'mountain' we would not be able to construct precisifications of 'mountain'.

Consider the dictum 'A precise conclusion cannot be deduced from vague premises'. This can be refuted. The disjunction of vague statement, V, and a precise statement, P, is either precise or vague. If 'P or V' is precise, then the counterexample is the argument: V therefore P or V. If 'P or V' is vague, then the counterexample is: P or V, not V, therefore, P.

Since precise conclusions can be deduced from vague premises, vague predicates can define precise predicates. For instance, 105 is precisely denoted by conjoining 'integer somewhat greater than 104' and 'integer somewhat less than 106' [Sorensen, 1988].

Vagueness does not collect over conjunction; a genus term can be precise where its species terms are vague. Consider the genus *Larus.* It contains ring species. In Britain, herring gulls are white and breed with the herring gulls of eastern America. American herring gulls are also white and breed with Alaskan gulls. Alaskan gulls breed with Siberian gulls. But as the specimens progress from Alaska and Siberia, they get smaller and have more black markings. When the series reaches back to Britain, they have become Lesser Black-Backed Gulls — which do not interbreed with herring gulls. Cases that are borderline between 'herring gull' and 'lesser black-backed gulls' are clear cases of *Larus.*

Vagueness also fails to collect under relationships that have resisted formalization. 'Color' is not the genus of 'red' and 'blue'. 'Red' and 'blue' are primitive terms; they cannot be defined as conjunctions of a genus and differentia. 'Red' and 'blue' are determinates of the determinable 'color' because they are mutually exclusive [Johnson, 1921, 176]. Cases that are borderline between red and blue are not borderline cases of color. Thus a precise determinable can have vague determinates.

The reverse holds as well: A determinable can be vague even though each of its determinates is precise. For there might be indeterminacy as to whether the determinates exhaust the range of the concept. Consider ‘number’. Each type of number is precise (natural numbers, negative numbers, rational numbers, irrational numbers, etc.) but there is still unclarity about whether there might be further types of numbers.

Vagueness also fails to collect meta-linguistically. There are vague predicates that denote all and only precise predicates. For instance, the vague predicate ‘precise predicate’ picks out all and only precise predicates.

The failure of vagueness to collect spells success in applying logic to vague predicates. That success is manifested in several ways that go beyond toy definitions.

Historically, scientific predicates develop from ordinary language. Just as ordinary language disambiguates itself, ordinary language precisifies itself.

Developmentally, vagueness precedes precision. The child’s innate stock of concepts is mostly vague, so even if a completely precise language were devised, children would not speak it. Parents need to communicate with their children, so the children would “corrupt” their parents. Only in the last stages of linguistic development do they master words for numbers and geometrical figures.

Linguistically, the applicability of logic to ordinary language is presupposed by the productivity of language. Instead of parroting back what parents said, children usually utter sentences that have never been spoken before. Children are systematic. Once they understand a sentence they understand all sentences of that sort. That understanding is reflected in their inferences. Children calculate the meaning of new sentences from the meaning of their components.

Logically, the topic neutrality of deduction ensures that classical logic applies to vague predicates. If an argument is formally valid, then uniform substitution of other predicates yields another valid argument. For what makes an argument formally valid is its structure, not its content. Therefore, if an argument with precise predicates is formally valid, uniform substitution with vague predicates will yield another valid argument.

6 DOES LOGIC ONLY APPLY LOOSELY?

According to Max Black, we can only apply geometrical principles to the extent that the objects do not significantly sag or bend. Logic presupposes sharp boundaries in the way geometry presupposes rigid boundaries.

What are the *specific* implications of this analogy for the sorites paradox? In “Reasoning with Loose Concepts” Black says that the sorites is valid but claims the induction step lacks a truth-value. This is a revision of standard logic rather than special handling of it. For Black is saying that an argument can be valid, have no false premises and yet have a false conclusion.

Black says the induction step lacks a truth-value because it quantifies over borderline cases. It “bundles together proper and improper instances” of the vague inductive predicate.

Black's bundling requirement is too strong. It rules out remarks about the relationship between clear cases and borderline cases. For instance, "If a man is clearly tall, then he is not a borderline case for 'tall"'. This is true, rather than neither true nor false. Indeed, it is just the sort of remark Black makes in the course of his analysis.

Perhaps Black could substitute another explanation of why the induction step lacks a truth-value. If he succeeds, he will still need to provide principles that govern inferences from statements that lack truth-values. So although Black only wished to restrict the application of classical logic, he is committed to the more formidable enterprise of devising a separate logic for vague statements.

7 TRUTH-VALUE GAPS

Might this logic be supervaluationism? Supervaluationists drop the requirement that all statements have truth-values. If a compound statement *would* have the same truth-value regardless of how truth-values were distributed to its components, then the statement actually has that invariant truth-value. This preserves all theorems of classical logic. The induction step of the sorites comes out false under all precisifications of the inductive predicate. Thus supervaluationism suggests an elegant solution to the sorites paradox.

Supervaluationism also harmonizes with the use theory of meaning. Words get their meaning from the intentions of speakers. These intentions are incomplete so our terms are not fully defined. As we go along, we fill in the meaning in accordance with new information and interests.

Supervaluationism differs from classical logic in not being truth-functional. One cannot calculate the truth-value of a compound statement from the truth-values of its components. It also requires rejection of inference rules such as contraposition, conditional proof and reductio ad absurdum. For now it is fallacious to infer that a statement is false from the premise that it is not true.

The meaning of the supervaluationist's rejection of the induction step is elusive. When the supervaluationist says 'There is an n such that it is safe to climb n meters up the mountain but not $n + 1$ meters', he warns us not to interpret him as agreeing with the epistemicist who thinks there is some specific number that is the last safe distance in meters. This seems like an ontologically dishonest use of the existential quantifier.

There is also obscurity in the supervaluationist's affirmation of the base step of the induction step. When he says 'It is safe to ascend one meter up the mountain', his logic construes this as the claim that the statement comes out true under all precisifications of 'safe'. But that seems different from calling something safe. The supervaluationist seems to be refusing to affirm that anything fits the vague predicate 'safe'. This makes supervaluationism resemble the nihilist solution that vague predicates do not apply to anything. For instance, Peter Unger [1979] denies that there is any ordinary thing but agrees that precise predicates apply to the world. He also agrees that vague predicates can be fruitfully applied (because of their relations with precise predicates). Thus there is less distance between the supervaluationists and the nihilist than first appears.

8 LOGICAL PLURALISM

Interest in a separate logic of vagueness goes beyond philosophers. The psycholinguist Stephen Pinker [1999] says we have a precise style of reasoning that is regulated by rules and a vague style of reasoning that is governed by similarity and memory. The precise style of reasoning is epitomized by Euclid's geometry. The vague style was first articulated as principles of association by the British empiricists. Ironically, the principles founded on introspection survived the behaviorist revolution and flourished as principles of conditioning.

Pinker uses irregular words to illustrate the strengths and weaknesses of the two cognitive subsystems. The rule for expressing the plural by adding the suffix $-s$ is a compact mechanism for generating indefinitely many nouns. However, the ending is sometimes hard on the tongue. Witness *sixths*. Or consider the dilemma over how to pluralize *sorites*. The rule says you need add the suffix $-s$. But this suffix is hard to hear in words that end in *s* so considerations of clarity militate against the s. You are damned if you do and damned if you don't. (Most writers prefer to be damned if they don't.) Irregular nouns are much easier on the grammatical superego; they use vowel changes (*goose, geese*), which are always easy on the ear.

The appropriateness of making exceptions is an issue in ethics. Generalists think morality is rule-based. Particularists think it is more akin to perception. There are extremists on both sides. Kant says you should never lie. Situation ethicists deny that there are any moral principles. The situationist has some wiggle room here because he could admit there are useful rules of thumb. Psychologists may contribute to this debate by gauging how well we perform when following rules and how well we perform opportunistically. Recent epistemologists are impressed by how simple rules (when based on actuarial reasoning) far outperform deliberation that seeks out the nuances of the situation [Bishop and Trout, 2005].

Physiologists may also have something to say. Mountaineering guides divide between the situationalists and the legalists [Boukreev, 1997, 155]. The legalists emphasize rules such as 'If you are not in sight of Mt. Everest's summit by noon, then turn back'. Situationalists say that such rules are not flexible enough and contradict the spirit of adventure. Legalism strengthens with altitude. The higher you go, the greater the effect of oxygen deprivation. The judgment of hypoxic individuals is impaired so they are best off following rules.

So what is the nature of language? Psycholinguists stress that speech is a reflex. There is no time to deliberate about your next phoneme. Speech, at the fine-grained level, must be automatic. That means rules will dominate. However, there is still room for memory and association to intercede when the performance of rules is weak.

There is much division of labor between rules and rote learning. There is also some competition in the overlap areas. Pinker conjectures that this duality underlies legal and social issues. In law, the drive for rules collides with the need to respect precedent and the quirks of human nature [Solan, 1999].

Epistemicists do not concede that classical logic has trouble with reasoning that relies on memory and association. An arbitrarily constituted set behaves just as logically as a set whose membership is dictated by a tidy rule. Some mathematical problems are solved by checking each case without any interest in subsuming the cases under a category. The logic of brute force is clear.

Like other human beings, logicians are slow to recognize randomness. They tend to read in more order than there is. But this fondness for order springs from our humanity not our logic. We like logic because we like order, not the other way around. Our logic must be powerful enough to represent our defeats as well as our victories. (The simple logical operation of negation goes a long way here.) So it has no problem with randomness and heart-breaking deviations from simplicity.

Epistemicists agree that there should be additional logics such as modal logic, epistemic logic, and mereology. After all, many valid arguments are not validated by classical logic. But a failure to validate is not the same as a verdict of invalidity. Classical logic is too weak in that it fails to validate arguments that owe their validity to terms conveying necessity, belief, part-whole relations, and other terms that are, arguably, logical constants. To correct this omission, logicians have developed supplemental logics that add inference rules or theorems.

However, the sorites argument raises the graver issue that classical logic commits an error of commission. Instead of saying nothing, classical logic validates sorites arguments. If the sorites argument is not valid, classical logic is too strong. The only cure would be a deviant logic, one that rejects a theorem or some inference rules of classical logic.

9 DEVIANT LOGIC

Instead of changing words to fit logic, the deviant logician changes logic to fit words. Most deviant logicians think the classical logician is a victim of his own success. Classical logic performs well in the arena (mathematical proof) for which it was designed. (The founding intuitionists are exceptions: the constructivist Brouwer thinks classical logic is fine for ordinary language but inadequate for mathematics.)

Since the deviant logicians characterize classical logic as an overextension from a local victory, they design their logics to agree with classical logic for the special case in which predicates are precise. Deviant logicians will say that their logic stands to classical logic as Einstein's physics stands to Newton's.

But this stills the leaves the question of which deviant logic is the right one for ordinary language. To say that a logic is right is to say that it accurately models the entailment relation. Considered narrowly, a logic is just a formal system in which strings of symbols are transformed into other strings of symbols. A formal system is neither right nor wrong. *Interpretations* of that system can be right or wrong. Interpretation takes us beyond formal logic into semantics, psychology, and metaphysics. That is where the philosophical action is and that is where I will focus.

10 MANY-VALUED LOGIC

Jan Łukasciewicz developed the first deviant logic. He proposed many-valued logic as a solution to Aristotle's problem of logical fatalism. According to the law of bivalence, every proposition has one of two truth-values, true or false. That includes propositions about the future such as 'There will be a sea-battle tomorrow'. If the proposition is true now, then the sea-battle is inevitable. If the proposition is false now, then the sea-battle is impossible. Thus the law of bivalence seems to imply fatalism.

Łukasiewicz's solution is that future contingent propositions only have a degree of truth. He gives truth-functional rules for calculating the truth-values of compound statements from the truth-values of their component statements. The truth-value of 'Not p' equals one minus the truth-value of p. Disjunctions get the highest truth-value possessed by a disjunct. So when p has an intermediate truth-value, 'p or not p' has an intermediate truth-value. The law of excluded middle is not a theorem. In particular 'Either there will be a sea battle tomorrow or not' is not fully true, so the fatalist loses a key premise.

Łukasiewicz's solution has been criticized. For instance, he says conjunctions get the same value as their lowest conjunct. So when p has an intermediate value, then 'p and not p' gets an intermediate value. Arthur Prior [1967] says this is too lenient. All contradictions should be false.

Such problems led to the invention of other variations of many-valued logic. Some add a fourth truth-value. Some add a fifth. Others add infinitely many. Some tinker with the rules for calculating truth-values. Some overhaul the rules wholesale [Edgington, 1996]. Others drop the truth-functionality altogether to preserve the theorems [Sanford, 1975].

Łukasiewicz's application of many-valued logic to fatalism was compromised by the strong performance of modal logic. The argument for fatalism in Aristotle's *De Interpretatione* appears to commit syntactic fallacies with the necessity operator such as inferring 'Necessarily p or necessarily not p' from 'Necessarily p or not p'. When a supplemental logic can cure the disorder we forego the deviant logician's surgery.

J. Goguen [1969] suggested that many-valued logic has stronger prospects as a logic of vagueness. He interprets borderline propositions as partially true. Goguen's rules suggest that the sorites argument is not perfectly valid. Although each conditional in the chain of reasoning is nearly fully true, they are not fully true. These departures from full truth are insignificant for short chains of reasoning. That is why classical logic looks fine. Indeed, classical logic is a special case of fuzzy logic where all the truth-values equal full truth or full falsehood.

11 FAMILY RESEMBLANCE

Those who say fuzzy logic is more applicable to ordinary language than classical logic often appeal to Ludwig Wittgenstein's doctrine of family resemblance. In Plato's dialogues, Socrates takes the ability to define to be the criterion of knowledge. To know something is to grasp its form. That form is expressed by a definition. A definition states a property that is common and peculiar to the concept in question. For instance, a geometer knows what a triangle is because he can define 'triangle' as an enclosed figure composed of three straight lines.

Although Socrates questions much, he does not question his presupposition that each word has a condition that is both necessary and sufficient for its application. Wittgenstein objects that many important terms (*number*, *art, consciousness, knowledge, religion*) instead get their unity from a family resemblance between cases. He illustrates with 'game':

> Look for example at board-games, with their multifarious relationships. Now pass to card games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost... is there always winning and losing, or competition between players? Think of patience [i.e., solitaire, in which there is no competition]. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared... Think now of ring-a-roses; here there is the element of amusement, but how many other characteristic features have disappeared!... And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing; sometimes overall similarities, sometimes similarities in detail. [sect. 66]

The positive aspect of Wittgenstein's proposal is that family resemblance is a sufficient condition for extending a term. However, Wittgenstein does not say which kinds of similarity are family resemblances. The fortuneteller examining tarot cards resembles the solitaire player but fortune telling is not a game. To ordinary people, dolphins seem similar to fish. To biologists, dolphins seem more similar to mammals. Here we relativize to experts. Should we go on to relativize to botanists and classify tomatoes as fruits? Some similarities are only heuristic (using gray hair to identify grandmothers) and are devoid of semantic import. Which are the semantically significant resemblances?

Most attention has been concentrated on a negative aspect of the family resemblance proposal: the absence of essential features. This challenge is not unprecedented. Medieval nominalists attacked realism about universals with a theory of analogy. W. E. Johnson [1921] contrasts the determinables/determinates distinction with the genus/differentia distinction by emphasizing that the former does not involved shared properties.

The positive side of the family resemblance doctrine is sketchy. Renford Bambrough suggested that the details be filled in along the lines of the following model.

Suppose **O**1 to **O**5 are five objects and ABCDE are five properties:

O1	**O**2	**O**3	**O**4	**O**5
ABCD	ABCE	ABDE	ACDE	BCDE

Each object shares 75% of its properties with each other object but none of the listed properties is common between all the objects.

Other passages from the *Philosophical Investigations* associate family resemblance with the continuity of a slippery slope.

> And we extend our concept of number as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres. (PI 67)

Each bundle of properties changes like the Ship of Theseus. All common elements are gradually lost:

O6	**O**7	**O**8	**O**9	**O**10
ABCD	ABCE	ABEF	AEFG	EFGH

Each object has a 75% resemblance with its immediate neighbors but not all of its neighbors. For instance, objects **O**6 and **O**10 have nothing in common. They are instead linked by their intermediate cases.

There certainly are many vague transitions in language. Lewis Carroll [1880, 31] has a word ladder in which he changes GNAT to BITE in six steps: GNAT, GOAT, BOAT, BOLT, BOLE, BILE, BITE. (Fred Madden cites this word ladder to explain why *Through the Looking Glass* has a gnat in a railway carriage alongside a goat.) Historical linguists document some linguistic changes as slippery slope transformations. Semanticists have noted the existence of sense spectra (a counterpart to ring species) in which a word changes its sense imperceptibly.

Wittgenstein admits that there are sufficient conditions for applying ‘game’; all team sports are games. He also admits that there are necessary conditions; all games have rules. At most, Wittgenstein is denying that there is a condition that is both necessary and sufficient.

But Wittgenstein does not quite deny that either. From a logical point of view, there is always something common and peculiar to any set of things. In Bambrough’s quota model, each object satisfies the following disjunction: ABCD or ABCE or ABDE or ACDE or BCDE. We can shorten the condition to $(A \& B) \vee (A \& C) \vee (B \& C)$. The common property can be expressed as a quota: the object must have 80% of the following properties: ABCDE.

When the extension is finite, we can state the necessary and sufficient condition as a long disjunction draw from the members of the predicate’s extension. Often we can devise a rule that compresses the statement of this essential property. But not always. If the shortest rule used to summarize the items is as long as the list, then we might as well rely on the list.

The vast majority of numerical sequences are not compressible. In the terminology of computer scientists, they have high algorithmic complexity and so are random. However, the limits of human memory ensure that speakers do not use predicates with large, uncompressible extensions.

Wittgenstein dismisses the list technique as a verbal trick: "One might as well say: 'Something runs through the whole thread — namely the continuous overlapping of those fibres" (PI 67). The logician is being more clever than wise because it is implausible that people follow such complex, abstract rules. Thus Wittgenstein is only denying that our usage is *explained* by condition that is both necessary and sufficient for our usage.

Socrates would accept the requirement that the necessary and sufficient condition be explanatory. For he believes we are guided by the definition of 'virtue' in the way a geometer is guided by the definition of 'triangle'. He is committed to more than the mere existence of the necessary and sufficient condition. Socrates thinks the knower must have the right psychological connection to the definition.

The nature of this connection might be more complicated than Socrates anticipated. Children rely heavily on a variety of heuristics (height, density, volume, item size) to judge whether one group has more items than another group: This makes them vulnerable to numeroisty illusions. As children mature, they learn to ignore these features when the task requires abstract reasoning.

Even a mathematician who knows the definition of 'greater than' might use memory or heuristics to judge whether there is more. People answer questions by applying a number of techniques simultaneously. When one of the techniques yields an answer, they stop. Thus, there is a race between methods. The mixed nature of the winners ensures psychological heterogeneity in the data. The bridge between the definition and human usage is paved with memory, encyclopedic knowledge, association, and rules of thumb.

Wittgenstein's second reply to the clever logician is that family resemblance terms are typically open-textured. Speakers are constantly adding to the meaning of the word. Think of all the different kinds of numbers (negative numbers, irrational numbers, etc.) that have been added to the extension — so far. This open frontier is part of the meaning of 'number'. Since there is no stable set of things to which the term applies, the logician cannot perform his trick of disjoining all the members of the extension.

Speakers do not always agree about whether to extend a term to a novel case. There is often no fact of the matter. Competent speakers can disagree.

This "faultless disagreement" [Wright, 2001] creates an interesting tension with Wittgenstein's theme that language depends on agreement. To be speakers of the same language, we must apply the same rules and judge the world in a similar enough fashion to give a point to the execution of those rules.

Socrates treats his failure to find definitions as a failure of ingenuity. The fact that one has not discovered the definition does not show that it does not exist. Admittedly, the *argumentum ad ignorantiam* seems less fallacious if the search has gone on fruitlessly since 399 BC. But the *Meno*'s doctrine of reminiscence suggests the outline of a rejoinder: Speakers have an impoverished knowledge of definitions. They are like students who can recognize the correct answers to a multiple-choice question but who cannot spontaneously answer (and so dread an essay exam). This impoverished knowledge is a stepping-stone to the richer sort of knowledge that goes with the explicit statement of definitions.

12 PROTOTYPES

The doctrine of family resemblance inspired Eleanor Rosch's [1978] research on prototypes. She reports that people perceive category membership as a matter of degree rather than a clear-cut yes-no matter. Reaction time studies show that people quickly and reliably judge central members. Flies are definitely insects. Robins are definitely birds. Their judgments are stable about whether cancer is a disease or whether apples are fruits. The story is different with peripheral members of categories. People are slow to judge whether a stroke is a disease or pumpkins are fruits. When re-tested a month later, many change their minds.

In fuzzy set theory, membership comes in degrees. Thus the fuzzy logician seems poised to consolidate Rosch's research. For instance, George Lakoff [1974] uses fuzzy logic to model predicate modifiers such as *somewhat*, *sort of*, and *loosely speaking*.

Fuzzy logic was given a further boost in the 1980s by the connectionist revolution in cognitive science. Classical systems of information are inspired by formal languages and the digital computers that instantiate those formal systems. They manipulate symbols by means of explicit rules in a serial manner. Connectionist systems are inspired by the brain. Their basic building block is the neuron. Connectionist systems are neural nets that rely on parallel processing of sub-symbols. They use statistical properties instead of all-or-nothing rules to transform information.

But the epistemicist acknowledges that statistical generalizations are often better than universal generalizations. We associate classical logic with universal quantifiers because simple syllogisms use universal quantifiers. But even Aristotle is careful to stress the role of the quantifier 'some'. There are puzzles about how to model quantifiers such as 'most'. The problem here is the *vagueness* of the quantifier, not its statistical aspect.

The last line of Peter Strawson's "On Referring" was intended as rebuke to Bertrand Russell's theory of definite descriptions: "Neither Aristotelian nor Russellian rules give the exact logic of any expression of ordinary language; for ordinary language has no exact logic." [Strawson, 1971, 27]. Russell later said this was their only point of agreement! But what would they be agreeing to? There are several ways to interpret Strawson's remark. Strawson might be saying that classical logic does not apply to ordinary language. He might be saying it applies only in some loose fashion or that a special logic is needed for vague language. Or Strawson might be saying that a deviant logic is needed. Our investigation has produced reasons to reject all three interpretations. Thus we are justified in continuing to apply classical logic to vague propositions.

BIBLIOGRAPHY

[Bishop and Trout, 2005] M. A. Bishop and J.D. Trout. *Epistemology and the Psychology of Human Judgment* New York: Oxford University Press, 2005.

[Boukreev and Dewalt, 1997] A. Boukreev and G. Weston Dewalt. *The Climb: Tragic Ambitions on Everest* New York: St. Martin's Press.

[Carnap, 1950] R. Carnap. *Logical Foundations of Probability*, Chicago: University of Chicago Press, 1950.

[Carroll, 1880] L. Carroll. *Doublets: A Word Puzzle* New York: Macmillan, third edition. 1880.

[Chesterton, 1927] G. K. Chesterton. The paradoxes of christianity. In his *Orthodoxy* New York: Dodd, Mean and Company, 1927.

[Edgington, 1996] D. Edgington. Vagueness by degrees. In *Vagueness: A Reader* ed. Rosanna Keefe and Peter Smith (Cambridge: MIT Press): 294–316, 1996.

[Fine, 1975] K. Fine. Vagueness, truth, and logic *Synthese* 30: 265–300, 1975.

[Gardner, 1989] M. Gardner. Penrose tiling. In his *Penrose Tiles to Trapdoor Cipers* New York: W. H. Freeman, 1989.

[Goguen, 1969] J. Goguen. The logic of inexact concepts, *Synthese*, 19: 325-73, 1969.

[Haack, 1974] S. Haack. *Deviant Logic* London: Cambridge University Press, 1974.

[Johnson, 1921] W. E. Johnson. *Logic*, Part I, Cambridge: Cambridge U. P, 1921.

[Nietzsche, 1878] F. Nietzsche. *Human, All-Too-Human,* 1878. Trans. Helen Zimmern. (New York: Russell & Russell Inc., 1964.

[Prior, 1967] A. Prior. *Past, Present and Future* Oxford: Clarendon Press, 1967.

[Russell, 1923] B. Russell. Vagueness *Australasian Journal of Philosophy* 1: 297–314, 1923.

[Sanford, 1975] D. Sanford. Borderline logic *American Philosophical Quarterly* 12/1: 29–39, 1975.

[Solan, 1999] L. Solan. Why laws work pretty well, but not great: words and rules in legal interpretation, *Law & Social Inquiry* 26: 243–270, 1999.

[Sorensen, 1987] R. Sorensen. The vagueness of knowledge *Canadian Journal of Philosophy* 17/4: 767–804, 1987.

[Sorensen, 1988] R. Sorensen. Precisification by means of vague predicates *Notre Dame Journal of Formal Logic* 29/2: 267–275, 1988.

[Sorensen, 1991] R. Sorensen. Vagueness within the language of thought *Philosophical Quarterly* 41/165: 389–413, 1991.

[Sorensen, 1998] R. Sorensen. Ambiguity, discretion, and the sorites *Monist* 81/2 (1998): 217–235, 1998.

[Sorensen, 2001] R. Sorensen. *Vagueness and Contradiction* Oxford: Clarendon Press, 2001

[Strawson, 1971] P. Strawson. On referring. In his *Logico-Linguistic Papers* Methuen: 1–27, 1971.

[Unger, 1979] P. Unger. There are no ordinary things *Synthese*, 4: 117–54, 1979.

[Williamson, 1994] T. Williamson. *Vagueness*, London: Routledge, 1994.

[Wright, 2001] C. Wright. On being in a quandary *Mind* 110/437: 45–98, 2001.

LOGIC AND SEMANTIC ANALYSIS

Ernest Lepore and Matthew Stone

1 INTRODUCTION

When we wish to frame or to communicate a precise and nuanced argument, we should first clarify whatever meaningful distinctions our reasoning exploits. That's why every good paper begins by defining its terms. A tiger is a large and ferocious predatory cat, yellow with black stripes. A bachelor is an unmarried man. Freedom is the capacity to choose one's actions for oneself, independent of causal forces in the outside world. Knowledge is justified true belief. Getting clear on our concepts is the process of analysis. It is such a fundamental part of philosophical practice that the preponderance of contemporary philosophical writing in English today is described as 'analytic'.

Analysis issues in statements that seem to equate ideas. When we agree on these statements, we seem to do so in advance of any argument. When we dispute them, we seem to argue not about what's true or false, but about what's worth discussing at all. These observations naturally lead to a particular interpretation of philosophical practice: When we do analysis, we ask ourselves about the meanings of our sentences or the contents of our thoughts. The meanings and contents we discover are inherent in our sentences and thoughts, and could not have been otherwise. And our introspection comes with the subjective authority with which we know our own minds. So when we practice analysis, the statements we derive are 'analytic'. Traditionally, this has been understood to mean that they are necessarily true, purely in virtue of the meanings of our terms, and that they are knowable *a priori*, independently of any experience.

If there were any analytic statements in *this* sense, it would be a boon for philosophy. We could argue against a skeptic about the real world, for example, by constructing an incontrovertible analysis of our concepts and reasoning from it to derive the knowledge we hope to attribute to ourselves from the incorrigible evidence of our senses. We could also argue against a skeptic about intersubjectivity — someone who contested the idea of public truth or falsehood on the grounds that our ideas are private and even language does not offer public means for us to share them. If we share the terms with which we frame our arguments, and know their meanings *a priori*, we can be confident that we agree on the questions, and so can be confident that our disputes contest the answers.

We shall follow Quine [1953] in thinking that there are no analytic statements in this broad sense. When we do semantic analysis, we cannot hope to infer

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

statements that have a special status in philosophy. We explicate his reasoning in this chapter, and suggest that while developments in the philosophy of mind and language over the more than fifty years since Quine may have superceded much of his perspective, they continue to support his conclusion. However, although we agree that there are no *a priori* truths that describe linguistic meaning, we do not believe semantic analysis should be abandoned. On the contrary, the enterprise of semantics in linguistics and in the philosophy of language depends on our ability to articulate and formalize precise statements that specify and explicate linguistic meaning. Such statements serve as crucial hypotheses in the scientific enterprise of accounting for linguistic behavior. Only such general laws of meaning can explain the productive capacity that we all have in order to use and to understand the potential infinity of statements in our native language.

Our view reflects the belief that the laws of meaning are not knowable *a priori*. As speakers, we learn them or fix them in our interactions with our linguistic community and the environment; as theorists, we hypothesize them to explain our linguistic competence. Thus, it is not a *logical* consequence of our knowledge of meaning that we share the meanings of our terms with others in our community; it is an assumption that we make but that a skeptic might deny. Moreover, it turns out that the semantic facts we need to explain language use are profoundly boring for philosophy more generally. Here is a typical example: If Kim kissed Sandy, then there was a kissing. Such facts do not bridge sensory and categorical vocabulary. They will not allow us to relate our knowledge of the world to the evidence of our senses; so they will not underpin our epistemology. Analysis, again, has clear limits, and we will have to fight the skeptic the hard way. Yet, all in all, we offer an optimistic view about analysis. If you already believe we have a broadly shared language to discuss an accurately perceived world, then semantic analysis points to a good explanation for how that could be.

Since Quine [1960], and especially since Montague [1974], formalism has become central to the practice of semantic analysis. Logic offers us an appealing, lightweight notation while nevertheless forcing us to be completely explicit and completely consistent in the ontology and concepts we appeal to in theorizing. When our arguments become difficult, this discipline becomes indispensable. The use of logic also anticipates the place for semantic analysis in a cognitive science of language and the world mediated by mental representations. Semantic analysis interprets linguistic meaning in terms of something fundamentally nonlinguistic: relationships in the real world. Here the nonlinguistic metalanguage of formal logic serves as a useful reminder that we have moved outside the realm of grammatical knowledge and communicative convention, into a system that holds its meaning through standards and mechanisms fundamentally different to those of language. In this chapter, then, we will also emphasize how formalism can be exploited (especially first-order logic but also higher-order logic and intensional logic), both to craft particular analyses more precisely and to think more generally about what analysis is.

2 SOME CLASSIC ARGUMENTS.

We begin with a review of three philosophical perspectives on knowledge of meaning. First, we present Quine's view of meaning as a theoretical interpretation of a speaker's present dispositions to assent to, or to dissent from, utterances in context. Quine's empiricism helps to make clear the intuition that there are no *a priori* truths about meaning. In fact, Quine argues that a theoretical interpretation is unable in principle to distinguish between facts about meaning and other empirical truths. Then, we discuss a different view of meaning, particularly one associated with Kripke and Putnam. Here the project of semantics is not to interpret a speaker's current behavior but to explain it; this kind of semantics will show how a speaker's thoughts and utterances can be meaningfully grounded in his prior experience, particularly his interactions with members of his community. This view predicts that there are truths about meaning: we can find statements that, just because of the meanings of the terms they contain, could not have been false. But on this view there still are no *a priori* truths about meaning, for truths about meaning are knowable only *a posteriori*. Finally, we link these considerations to the architecture of contemporary semantic theory, drawing on the interpretive truth-theories of Davidson [1967b; 2001], as fleshed out, for example, in [Larson and Segal, 1995]. Linguistic methodology challenges us to explain the systematicity and productivity of meaning in language. One response to this challenge follows Kripke and Putnam, but of course not Quine, in appealing to explicit knowledge of meaning. On this account, we formulate novel sentences and recognize the meanings of others' novel sentences by reasoning directly from facts we have learned about the meanings of expressions of our language — and our mental apparatus stores these semantic facts and computes them with a nonlinguistic formal system — 'a language of thought' [Fodor, 1975], for which logic provides an attractive model.

2.1 Quine on analysis

Quine [1960] developed his arguments about analysis as part of a larger project to make sense of commonsense mentalistic descriptions of behavior within a rigorous, scientific framework. Like many philosophers before and since, Quine was deeply skeptical about the constructs and explanations of commonsense psychology. Indeed, in his philosophy, Quine generally regarded all knowledge as incomplete, provisional, and open to revision. To describe an agent's beliefs, Quine dispensed with theoretical constructs such as mental representations and abstract objects such as meanings and propositions. He focused instead on characterizing the sentences which an agent will assent to and the circumstances in which the agent will do so. In this way, Quine sought to refer mentalistic descriptions straightforwardly to observable distinctions; our psychological theories could then exploit these distinctions to interpret the statements and actions of others against our own theoretical picture of the world.

Against this background, Quine argued that meaning is indeterminate. There could be many, equally valid schemes to render the content of another's beliefs in our own terms. Each of these schemes would be holistic, in that it would aim to provide a consistent overall interpretation of a body of beliefs in terms of our own. The key to such an interpretation would be its overall coherence; as far as possible, it would have to interpret true sentences as true and false sentences as false. To judge another to be systematically in error would undercut our claim to understand his sentences. Quine saw this holism of interpretation in mental explanation as continuous with the holism of other kinds of scientific theorizing. For Quine, talk of hypothesized entities in science, like electrons, derives its content from its role in a network of principles that ultimately leads to empirical predictions.

Quine's famous 'Gavagai' thought experiment illustrates both his holistic view of meaning and his empirical, *a posteriori*, understanding of the attribution of semantic content. He invites us to consider a field linguist working with speakers of an isolated language without access to any interpreters. The task of this linguist is to arrive at translations from the speaker's language into his own for heard utterances. Quine calls this project 'radical translation'. Of course, the linguist must consult speakers as informants, so she first discovers how they signal agreement and disagreement. Now to get a handle on what a particular utterance means, she determines the situations that prompt informants' agreement with it, and those that prompt their disagreement with it. She settles on a good translation by finding a statement in her own language for which her judgments match those of her informants. She discovers, for example, that 'Gavagai!' can be translated as 'Look, a rabbit!' by seeing that informants accept 'Gavagai!' in more or less the situations where she would accept 'Look, a rabbit!' — in short, when there is a rabbit around.

Can translations so discovered be regarded purely as facts about meaning? Quine argues they cannot. The linguist's procedure yields translations that inevitably summarize and reflect a broad swath of informants' empirical knowledge. An informant's assent to 'Gavagai!' will be grounded perceptually, in his own experience of the appearance of rabbits and the contexts where rabbits are found, and it may also be grounded socially in information provided by his companions or simply in the shared background of his broader culture. In arriving at a translation, the linguist *must* locate a sentence in her own language that approximates the *entire* network of factors which determine her informant's judgment. She cannot narrow the scope of translation by ruling out some of these factors as extrinsic to meaning. After all, she aims to *discover* her informant's meanings; if she rules out potential aspects of meaning in advance, she abandons science for mere prejudice. Moreover, in construing some information as extrinsic to meaning, she worsens the fit between her informant's judgments about his utterances and her judgments about the translations. In this sense, she really does less faithfully capture what her informant means.

For Quine, a child learning her native language is in the same situation as a scientist translating an exotic new language. She must construct a radical interpretation of the utterances of her community. Her conclusions take the form of an overall theory about how people act in her community, and this theory must be assessed holistically against all her empirical evidence. The theory must allow her to understand the sentences members of her community accept as true, and those members of her community reject as false. When she evaluates her interpretation this way, she can make no principled distinction between what 'rabbit' *means* in her community, and what rabbits *are* to her community. If members of her community had assented to other sentences involving the word 'rabbit', her theory would have done better to interpret 'rabbit' differently. This is why Quine suggests that a theoretical interpretation is unable in principle to distinguish between facts about meaning and other empirical truths. No sentences are true solely in virtue of meaning. A fortiori, there can be no 'analytic' sentences that are true solely in virtue of meaning, that are necessarily true, and that can be known *a priori.*

2.2 Kripke, Putnam, and the necessary a posteriori

Quine conceived of interpretation as our effort to *understand* one another. Interpretation is charitable. In constructing an interpretation, we primarily aim to find points of agreement between our perspective and another's. A different project in semantics is to try to *explain* one another. An explanation gives meaning to another person's thoughts and sentences by describing his relationship to the world around him. Explanation need not be as charitable as interpretation. Indeed, the appeal of explanatory semantics springs from thought experiments where we are prepared to judge another to be systematically in error, and yet we refrain from reinterpreting his sentences.

Here is a clear case. With Putnam [1975], imagine a planet exactly like Earth, except that where Earth has water, this other planet, Twin Earth, has another, mysterious substance, XYZ. To human senses, this substance seems exactly the same as water; nevertheless, it has a fundamentally different chemical structure. Imagine further that it's still the year 1700, and chemical structure has yet to be discovered. Still, we judge that the English word 'water', on Earth, means water, whereas the Twin English word 'water', on Twin Earth, means XYZ. Moreover, if an Earthling were suddenly teleported to Twin Earth, he would still speak English, and his word 'water' would still mean water — this despite the fact that he might have exactly the same dispositions as Twin Earthers have to accept or reject statements about his new surroundings. In short, the unfortunate Earthling would think he was surrounded by lots of water, and he would be completely wrong.

You can, perhaps, take Putnam's Twin Earth thought experiments to illustrate why Quine thought translation was indeterminate. When you look at how a speaker is disposed to respond to sentences of English, you can interpret 'water' equally well as water, XYZ, or the disjunction of the two. These interpretations are different, and they would assign different truth values to English statements in meaningful (but ultimately inaccessible) situations. In fact, though, when we say that 'water' in English means water, we are applying a different standard than Quine's. Kripke's [1972; 1980] suggestion is that our standard is based on our understanding that English speakers *intend* to pick out a particular kind of stuff in their own environment. As a community, English speakers have encountered this stuff and named it 'water'. And as a community, English speakers work together to ensure first that the community maintains the referential connection between the word 'water' and that stuff, and only secondarily that individuals in the community can themselves recognize examples of the stuff in particular situations. When as observers we recognize that 'water' means water, we aren't summarizing the epistemic abilities of particular speakers. Rather, we are summarizing social commitments and causal connections in the community that have worked across speakers to hook the word 'water' up with the stuff, and keep it that way. What we do, ultimately, is to *explain* how speakers can use language to refer in shared ways to aspects of the world.

Kripke motivates his view with an analogy between words for kinds, such as 'water', and proper names, such as 'Richard Feynman'. In the case of proper names, we can point to the social practices that initially fix the reference of a name and transmit that reference within the community.

> Someone, let's say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say, Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman is a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker. He is then referring to Feynman even though he can't identify him uniquely. He doesn't know what a Feynman diagram is, he doesn't know what the Feynman theory of pair production and annihilation is. Not only that: he'd have trouble distinguishing between Gell-Mann and Feynman. [Kripke, 1980, 91]

The result is that we can judge a speaker's reference with a proper name independent of the sentences that the speaker would assent to or reject.

In the case of common nouns like 'water', the word has had its reference since time immemorial. Nevertheless, new speakers still link themselves into chains of reference that participate in and preserve the connection between 'water' and water. So analogously, we take an English speaker's word 'water' to refer to water, independent of the sentences that the speaker would accept or reject.

Most philosophers find this new view of semantics more satisfying than Quine's in key respects. The new view offers a close fit to our intuitive understanding of ourselves. It seems that we really do commit to use our words with the same reference as our community. And when others make claims about the world, it seems that we really do assess and dispute those claims with respect to the common standard in the community, not just with respect to our interpretation of others' personal theories of the world. Moreover, the new view suggests specific grounds to critique the motivations behind the Quinean interpretation. For example, on this new view, we will inevitably focus on certain aspects of an agent's verbal behavior and not others when we assign meanings to his utterances. We do so because we locate semantics as *part* of a broader science of the mind, which combines a theory of language with a theory of action (including an account of our intentions and social relationships) and a theory of perception (including an account of the limits and failings of our observation). Semantics in itself explains only so much — and, not surprisingly, just because we understand the meaning of someone's sentences, we do not *ipso facto* understand *them*.

Crucially, this new view of semantics predicts that some statements are necessarily true, solely in virtue of the meanings of the words involved. We've already seen that it is a fact about meaning that 'Richard Feynman' names Richard Feynman, or that 'water' names water. We can go further. 'Hesperus' names the planet Venus, 'Phosphorus' names the planet Venus, 'is' names the identity relation. So sentence (1) follows, just as a matter of meaning:

1. Hesperus is Phosphorus.

Given that 'Hesperus' and 'Phosphorus' are both names for the planet Venus, (1) has to be true. There is no way that *that planet* could have failed to be *that planet*. Like sentence (1), the other facts that follow from the meanings of our language are necessarily true.

However, these facts about meaning are not knowable *a priori*. We discovered them. To illustrate the point, imagine that, early on, the Ancient Greeks were in an epistemic situation that left it open whether the bright object that sometimes appeared in the morning sky was the same as the bright thing that sometimes appeared in the evening sky. They could not distinguish themselves from their doubles on a Twin Earth where the morning star and the evening star actually were distinct objects (alien satellites, we might suppose). These Twin Earthers would speak a language in which (1) translates into a false sentence — indeed, a necessarily false sentence. For the Ancient Greeks, however, the translation of (1) was necessarily true. Eventually, the Ancient Greeks advanced their science, and improved their epistemic situation. They realized that in our case there is only one celestial object. At the same time, then, they discovered that (1) is necessarily true.

When we reflect on the generality of the Twin-Earth thought experiments, it is clear that semantic facts are knowable *a posteriori*. We can imagine being quite wrong about what our world is like. In these imaginary situations, our empirical errors extend to errors we make about what our words mean. And, of course, we can also imagine disagreeing with others about what the world is like. Though we are committed to use our words with the shared reference of our community, we must be prepared to resolve our dispute by relinquishing facts that we think are necessarily true — facts we think characterize the meanings of our words and the contents of our thoughts. Thus, the situation for philosophy generally remains much as Quine understood it.

Contra Quine, we can pursue semantic analysis, and attempt to discover statements that are necessarily true in virtue of the meanings of the terms. But any insights we derive about meaning are *a posteriori*. So, as Quine concluded, they are hypotheses that we make and test like other empirical statements; they do not have a special status in our arguments in virtue of the fact that they characterize meaning.

2.3 Knowledge of meaning and linguistic competence

The development of linguistic semantics as a subfield of cognitive science has confronted philosophical views of meaning with new evidence of the richness and sophistication of speakers' semantic judgments about their native language. To start, speakers are able to understand and produce indefinitely many sentences — sentences neither they nor anyone else in their community has uttered before. This shows their knowledge of language must be *productive*; it must go beyond a fixed lexicon of predefined static elements, and must include a generative system that actively composes linguistic knowledge so as to describe arbitrarily complex structures. The hallmark of productivity in language is recursion. Recursive patterns of complementation, as in (2), and recursive patterns of modification, as in (3) and (4), allow phrases to be nested indefinitely many times within a single sentence.

2. Chris thinks that Kim thought that Robin wanted Sandy to leave.

3. Chris bought a gorgeous new French 3-quart covered copper saucepan.

4. Chris is writing a book that describes inventors that have built machines that changed the world that we live in.

Speakers' capacity to formulate and recognize an open-ended array of possible sentences shows how acute a problem it is to coordinate meaning across speakers. When we learn the meaning of our native language, we must generalize from the finite record of our previous experience to an infinity of other sentences and situations. If we thereby arrive at a common understanding of the meanings of sentences, it must be because language is structured by substantive and inherent constraints that we are able to exploit. The same goes for theoretical projects in semantic analysis. We should expect that our discoveries about the meanings of sentences will generalize to an indefinite array of cases, and that they will respect and illuminate linguistic constraints. More generally, if our discoveries in semantics are to help explain how speakers can use language meaningfully, we should expect that the generative mechanisms we postulate as theorists will be compatible with the psychological mechanisms that underlie speakers' abilities.

A straightforward way to explain productivity in semantics is to assume that the meanings of particular sentences can be calculated by inference from general facts about meaning in the language. For example, consider the theory presented in examples (5)–(7).

5. 'snow' is a noun phrase and refers to the stuff snow.

6. 'white' is an adjective phrase and refers to the property whiteness.

7. If N is a noun phrase and refers to the stuff S and A is an adjective phrase and refers to the property P, then $\ulcorner N$ is $A \urcorner$ is a sentence and is true if and only if S is P.

From this theory, we can derive (8) as a logical consequence.

8. 'Snow is white' is true if and only if snow is white.

We can take (8) as a characterization of meaning, because it links up a sentence, in this case the sentence 'Snow is white', with a condition in the world stated in objective terms, in this case, the condition that snow is white. As semantic theorists, we can use this kind of theory, which Davidson [1967b; 2001] calls an interpretive truth-theory, to provide a general account of how sentences link up with conditions in the world. We use atomic formulas to axiomatize the semantics for the elementary structures in the language and use conditional formulas to describe the semantics of complex structures in the language as a function of the semantics of their constituents. We then reason logically from the axioms to associate particular sentences with conditions in the world. As in (5)–(7), this inference will be *compositional*, in that the conclusions we derive will be inferred through a logical derivation that mirrors the syntactic derivation of the sentence.

There are two ways to view interpretive truth-theories such as (5)–(7). We can use an interpretive truth-theory to formulate a semantic theory for a new language. For example, we could be pursuing radical translation. In this case, we are interested in systematically articulating translations for sentences in the object language in terms of sentences in our own; we understand these translations to be derived by inference from the axioms of the theory. This is compatible with a holistic, interpretive view of meaning like Quine's — indeed, Davidson himself subscribed to such a view.

Another way to view interpretive truth-theories such as (5)–(7) is as ingredients of speakers' psychology. On this view, we regard axioms such as (5)–(7) as generalizations that native speakers know tacitly about their language. When speakers formulate or recognize particular utterances, they reason tacitly from this implicit theory to derive conclusions such as (8), about specific new sentences. On this understanding, interpretive truth-theories offer an explanation of how speakers' knowledge of meaning and their inferential abilities together underlie their semantic competence.

This psychological understanding of interpretive truth-theories meshes better with an explanatory view about truth about meaning than with Quine's conceptual framework. For this psychological understanding commits us to a *representational theory of mind*, in which psychological processes involve the tokening of symbolic structures and these structures are locked on to content about the world through causal or informational mechanisms. (For more, see [Fodor, 1975] or [Fodor, 1987].) To see this, let us return to (8):

8. 'snow is white' is true if and only if snow is white.

As cognitive scientists, we expect that speaker knowledge of linguistic meaning will place words and sentences in correspondence with represented regularities in the world. In this case, the correspondence links the sentence 'snow is white' with an answer as to whether snow is in fact white: a distinction you have a handle on some other way. Theories of mental content that postulate causally-grounded nonlinguistic structured representations give you a way to make sense of this idea. A speaker's mental representation of the distinction whether snow is white is composed from mental symbols SNOW for the stuff snow and WHITE for the property of whiteness. The meaning of these symbols follows from their intrinsic role both in the speaker's mental life and in his interactions with his environment. Further symbols, again with a suitable interpretation, represent the property of truth (TRUE) and the logical biconditional ($\leftrightarrow$). On a representational theory of mind, these symbols can now be assembled into representation (9) corresponding in form and meaning to (8):

9. TRUE('snow is white') $\leftrightarrow$ WHITE(SNOW)

For an individual speaker, the tacit derivation of this symbolic structure can fix the meaning of the sentence 'snow is white'. That is, (9) represents a hypothesis about the meanings of sentences in the speaker's language that the speaker is committed to. At the same time, it shows how the speaker's sensitivity to the meaningful connection between language and the world can be founded directly on the natural intentionality of the speaker's concepts and thoughts.

The willingness to appeal to the representational theory of mind as part of a scientific reconstruction of commonsense psychological explanation is one of the signal departures from Quine that motivates analysis in contemporary philosophy. On Quine's behavioristic perspective, we capture mental content directly through a network of beliefs, circumstances and behaviors. We must assign meaning to agents sentences without recourse to intermediate representations. And that means statements such as (9) are not special; they are known to be true only as part of a holistic explanation of all the empirical evidence — a speaker does not know that they are true in virtue of meaning. For Quine, (9) can summarize the meaning of 'snow is white' but it cannot establish it. In this sense, the representational theory of mind is essential if we want to explain speakers' knowledge of meaning by appeal to their tacit knowledge of a theory of truth.

3 DOING SEMANTIC ANALYSIS BY ANALOGY TO LOGIC

In §2, we sketched some arguments for the position that semantic analysis is both possible and philosophically important in characterizing speakers' rule-governed knowledge of linguistic meaning. This view invites an analogy between the semantics of natural languages and the semantics of the artificial languages of formal logic. The analogy goes back to Frege [1879; 1967], who took logic to clarify the features of natural language essential for correct mathematical thought and communication. Davidson's [1967b; 2001] interpretive truth-theories, exhibited in (5)–(7), are built on Tarski's [1935; 1956] ideas about the semantics of formal languages. The work of Montague in the 1960s and 1970s took the analogy further [Montague, 1974]. Montague explicitly advocated an exact parallel between the semantic analysis of English — what ordinary speakers actually know about their language — and the semantics of intensional higher-order logic. In fact, many techniques originally developed for giving semantics to logical languages turn out to be extremely useful in carrying out semantic analysis.

We can briefly describe two notable examples that illustrate this. The first comes from the semantics of first-order logic, where Tarski [1935; 1956] introduced assignment functions to give semantics to occurrences of variables. The problem is to explain the truth conditions of complex logical formulas such as (10) in terms of their constituent logical formulas, such as (11). In effect, we have to answer what x means.

10. $\forall x F x$.

11. Fx.

Of course 'x' is a variable — it means whatever we want it to mean. Our semantics should simply make this precise. One way to do this is to use a *function* from the set of variables to the set of values—this function then represents a specific set of choices about what the variables correspond to. Such functions are called assignment functions; we use g to name a representative assignment function. Thus (11) is not simply true or false; it is true or false with respect to a set of choices about the variables. Specifically, (11) is true with respect to an assignment g just in case the individual picked out by $g(x)$ has the property represented by F. This framework allows us to understand (10) intuitively in terms of (11): if (11) comes out true for every choice we make for the value of x, then (10) is also true. More precisely, (10) is true with respect to an assignment g just in case (11) is true for any assignment g' exactly like g except possibly that $g'(x)$ differs from $g(x)$.

This treatment of variables and quantifiers provides a useful tool for approaching the analogous meanings of corresponding constructions in natural language, such as quantified noun phrases, bound pronouns and traces. For instance, consider sentence (12). Linguists understand (12) as composed through a syntactic derivation that makes the scope of the quantifier explicit but still indicates that the quantifier supplies the reference for the subject of the sentence and for the pronoun *her*.

12. Every student finished her homework.

The logical form (or LF) in (13) represents these relationships by one common convention: the quantifier is *indexed*, or associated with a variable i; it is *moved*, so that it brackets a sentence that indicates its scope (just like a logical quantifier); it *leaves a trace*, written as t_i, a variable in the subject position of *finished*; and the pronoun *her* is also indexed with the same variable i.

13. Every student$_i$ (t$_i$ finished her$_i$ homework).

14. t$_i$ finished her$_i$ homework.

This LF representation invites us to specify the semantics of (13) in terms of the semantics of (14), and the formalism of assignment functions originally developed for logical languages works beautifully for this. (13) and (14) are related just as (10) and (11) are. For a philosophically-accessible tutorial of the use of assignment functions and other logical techniques to interpret natural language quantifiers, see [Pietroski, 2006]. Analogies between pronouns and variables in formal languages continue to provide fertile grounds for semantic exploration. For example, dynamic semantics is an approach to the analysis of pronouns in natural language that manipulates assignment functions using formal tools originally developed to account for the meanings of variables in the formal languages used to specify computer programs; see for example [Groenendijk and Stokhof, 1990] or [Muskens, 1996].

A second arena where techniques originally developed in the semantics of formal languages provide instructive tools for natural language is in the treatment of time and possibility. Temporal and modal logics feature sentential operators that say when or in what hypothetical conditions a sentence is true. For example, in modal logic, the formula $\Diamond p$ says that p is possibly true. Here the semantic problem is to specify how a formula that says that p is possibly true can get its semantics in terms of a constituent formula p, which says that p is actually true. Again, we need to be more precise about the contribution that constituent formulas make to the truth conditions of larger formulas. In $\Diamond p$, we don't use p to say that the proposition is actually true; we only say that the proposition is true in a hypothetical possibility that is under consideration. We need to relativize our formal definition of truth accordingly.

Suppose then that we have a set of possible *points of evaluation*, corresponding to the different times, possible worlds, or other situations our logic describes. We use w as a variable ranging over these points of evaluation. The semantics for the logic will then describe whether a formula is true at a particular point of evaluation. Thus, for example, $\Diamond p$ is true at a point w if there is some other point w' that is possible from w and p is true at w'. Normally, we distinguish one of the possible points of evaluation as the present one. A formula is simply true if it is true at this present point. See [Fitting and Mendelsohn, 1998] for a philosophical presentation of the resulting logics.

Of course, natural languages have constructions that seem to depend on times and possibilities in ways analogous to the formal languages. We list a couple in (15) and (16).

15. It's possible that Sandy left.

16. Whenever it rains, Chris brings an umbrella.

To describe the contribution 'Sandy left' makes to (15), or 'Chris brings an umbrella' makes to (16), it is natural to appeal to a notion of truth at a point of evaluation that includes temporal and modal information. (15) is true at a point w if there is some point possible from w where 'Sandy left' is true. (16) is true at a point w if every temporal point comparable to w where 'it rains' is true is also one where 'Chris brings an umbrella' is true.

As in the case of pronouns, pursuing this analogy in detail leads to a rich and profitable interplay between logic and linguistics. When we explore natural language semantics systematically, we find that the tenses of natural languages have some of the properties of operators in temporal logic, but they also have some of the properties of pronouns [Partee, 1975]; [Hinrichs, 1985]; [Webber, 1986]. Modality in natural language is similar [Roberts, 1989]. Logically, this behavior finds a match in hybrid logics [Areces *et. al.*, 2001] — formal extensions of modal logic that allow formulas to mark the points visited in evaluating a formula and to return to those points elsewhere to describe them further.

While progress in formal semantics for natural language has often been supported by parallels between language and logic, it remains true that philosophical argument often requires us to appreciate the differences between everyday language and idealized uses of formulas in familiar logics. Language depends on context. Logic, by contrast, is so useful in part because it doesn't depend on context: we use logic to present our reasoning in a form whose correctness is intrinsic to its explicit structure and can be determined once and for all. As we explain in more detail in §§3.1-3.2 formal methods for the treatment of context-sensitivity in natural language complicate the picture familiar from logic. Accordingly, we need more nuanced distinctions to describe semantic relationships in languages with context-sensitivity than in languages without it.

Another important distinction is that logical formulas are regimented with specified semantics so that they wear their meanings on their sleeves. By contrast, as we emphasized in §2, natural languages must be learned. We can be uncertain about the meanings of sentences, and we can be creative in how we use them. As we explain in more detail in §3.2, formal accounts of the interpretation of natural language must take this flexibility into account. Again, we find that we need more nuanced distinctions in natural language to describe the relationship between the meaning of a sentence and the way we understand it than logical languages typically call for.

3.1 Indexicals, demonstratives, and context sensitivity[1]

English has two kinds of terms that are not typically included in the languages of formal logic: demonstratives and indexicals. Demonstratives are words like ‘that’, ‘there’, or ‘he’. Paradigms of indexicals are ‘I’, ‘now’, and ‘here’ (at least on one of its senses).[2] What a token of any of these terms refers to depends crucially on the context in which it is tokened. A token of a demonstrative word requires an associated *demonstration* to fix its referent. Typically, but not always, a demonstration is a visual presentation of a salient local object discriminated by a pointing gesture. The token of a demonstrative expression refers to whatever its accompanying demonstration demonstrates — what we shall call a *demonstratum.* Unlike a demonstrative, a token of an indexical requires no accompanying demonstration to secure its reference. The reference of an indexical is determined directly by information about the situation in which it is tokened.

Context sensitivity is a challenge to the analogy between the natural language and logic. As we shall see, context sensitivity requires us to depart from the formalism of logic; this is neither a surprise nor, in itself, a worry. More significantly, context sensitivity requires us to make distinctions in analyzing utterances that we do not need when we describe logical statements. These new distinctions are significant. Chief among them is the distinction between *linguistic meaning* and *utterance content.*

The *linguistic meaning* of an expression is what competent language users learn when they learn that expression. Learning the linguistic meaning of an indexical expression is learning how to adjust its reference across contexts of use. The linguistic meaning of ‘tomorrow’ determines what that word picks out across varying contexts of its use, as in the designation rule (T).

> (T) In each context of use, ‘tomorrow’ refers to the day after the day of that context of use.

The same goes for demonstratives. Anyone who comprehends the linguistic meaning of a demonstrative expression knows that a satisfactory use of it requires the speaker to make an accompanying demonstration, so, e.g., anyone who understands the demonstrative word ‘that’ must know its linguistic designation rule (D).

> (D) A token of ‘that’ refers to the object demonstrated by whoever uses that token.

Linguistic meaning underwrites translation. Suppose that Mike utters (17) yesterday:

17. The Yankees will win tomorrow.

---

[1] This section is largely a summary of the important work of Kaplan [1989].

[2] ‘here’ is a indexical in ‘I am in here’, but a demonstrative in ‘In two weeks I will be here’ (pointing at a city on a map).

(18) is the translation of (17) into Italian. The reference of ‘domani’ is fixed in Italian by a rule like (T) analogously to ‘tomorrow’ in English.

18. Gli Yankees vinceranno domani.

If Massimo utters (18) on the same day that Mike utters (17) then the two have said the same thing. Thus, in one sense, two sentences with the same linguistic meaning say the same thing.

However, there is another sense in which two sentences with the same linguistic meaning can say different things. *What you express* in this sense is the *content* of your utterance. Distinct utterances of ‘That’s a flower’ can disagree in truth-value because the objects demonstrated might be different. If your two tokens of ‘that’ demonstrate distinct objects, obviously you were not discussing the same object on both occasions. But if your topic of conversation changes, then what you say (or express) on these distinct occasions cannot remain the same. This might be so even though the distinct circumstances surrounding your two tokens might go unnoticed by you (and your audience). The lighting in the room, the sorts of gestures that you make, and every other aspect of your accompanying demonstrations might be indistinguishable to both you and your audience. Demonstratives are quite special linguistic devices in this way: we can token the same words with the same linguistic meaning in contexts that are indistinguishable to the speaker and auditor, and yet, unwittingly, say (or express) different things because different things are referenced.

We can also use different words with *distinct linguistic meanings* to say (or express) the same thing. Should we say, ‘He is a very nice person’ pointing at Bill, and you speaking to Bill say, ‘You are a very nice person’, then we agree in what we say about Bill, even though our words ‘you’ and ‘he’ differ in linguistic meaning. In sum, the same context-sensitive expression can be used on different occasions to say something different, and different context-sensitive expressions distinct in linguistic meaning can be used on the same (or different) occasions to say the same thing. These sorts of considerations recommend that we distinguish two sorts of meaning contributions demonstrative expressions can make.

This distinction has pervasive consequences for semantic analysis. We have to cash the distinction out formally, by specifying the meaning of utterance in terms of an occasion of use distinct from its circumstances of evaluation. And we have to follow the distinction conceptually, particularly when we think about whether what we say is *a priori* or metaphysically necessary.

Earlier we observed the usefulness of describing the temporal and modal interpretation of utterances by analogy to the interpretations of modal and temporal logics. Sentence (19) is a sentence in future tense without an (explicit) indexical noun.[3]

19. The earth *will be destroyed* by a large asteroid.

---

[3] We say ‘no explicit’, since tense itself introduces an indexical element.

An utterance of (19) is true at a time t just in case at some later time t' sentence (20), which is in present tense, is true.

20. The earth *is destroyed* by a large asteroid.

Sentence (6) is also in future tense, but unlike (19), (21) has in it an explicit context sensitive expression, namely, the word 'now'.

21. Everything *now* red will be blue.
    (or, 'It will be the case that everything *now* red is blue'.)

Were we to reason by analogy with (19), we would have to conclude that (21) is true at time t just in case its present tense counter-part (22) is true at some time t' later than t.

22. Everything now red is blue.

But that's impossible, since (22) is true at a time t' just in case everything that is red *at time* t' is also blue at t', and such circumstances are not physically possible. Now since an utterance of (21) can be true, it follows that the truth of (21) at a time t cannot require the truth of (22) at a time t', later than t.

Why does the analogy between (19) and (21) break down? It does so on account of the indexical expression 'now' in (21). In evaluating an utterance of (21), we simply cannot ignore that the word 'now', by virtue of its linguistic meaning alone, must pick out the time at which (21) is tokened (used, uttered): so (21) is true at a particular time t just in case (23), which lacks any context sensitive noun, is true at some time t' later than t,

23. Everything (that is) red *at* t is blue.

where, let's assume, 't' is a proper name that refers to a particular time t, namely, the time at which (21) was uttered. And these conditions seem to get things just right. An utterance of (21) is true at a time t just in case (23) is true at some later time t', i.e., just in case those red objects at t have changed color to blue at t'. The indexical expression 'now' tokened at t, by virtue of its linguistic meaning alone, must pick out the time of its utterance, in this case, t.

When evaluating an utterance of a sentence in future tense, obviously, the relevant circumstances for its evaluation (those circumstances that the utterance is about) are in the future. (21), by virtue of its future tense, is about circumstances after any of its utterances. Likewise, when evaluating a sentence in past tense, the relevant circumstances of evaluation for any of its utterances (those circumstances that utterance is about) are in the past; and so on.

What makes indexical expressions special is that the references of their tokens cling to the occasions on which they are tokened, and these occasions need not be the relevant circumstances for evaluating the truth or falsity of an utterance of sentences in which the indexicals occur. So even though a context in which someone utters (21) cannot by itself determine whether that utterance is true, that context and that context alone, and not, e.g., any future circumstances, suffices to determine the referent of its indexical expression 'now'.

Turning from tense to modality, the same sort of disanalogy arises for sentences about *non-actual* circumstances. Consider the modal sentence (24) with its context sensitive noun 'I'.

24. It is possible that *I* do not exist.
    (or, 'I might not have existed'.)

An utterance of (24) does not claim that its non-modal counterpart (25) can *possibly* be uttered truthfully.

25. I do not exist.

(25) can never be uttered truthfully. The linguistic meaning of 'I' is given by the designation rule (R).

> (R) In each context of use, a token of 'I' refers to the agent of that context.

So (25) is false if 'I' has a referent in that circumstance, because the speaker exists, and it lacks any truth value otherwise, because if 'I' fails to secure a referent. Therefore, if an utterance of (24) is true only if its non-modal counterpart (25) is true in a possible circumstance, then it follows that every utterance of (24) is false.[4] But our intuitions tell us otherwise; every utterance of (24) is true (assuming no necessary beings exist).

Again, we must distinguish between occasion of use and point of evaluation. For any utterance of (24), its token of 'I' picks out *the speaker of that utterance.* So, in evaluating an utterance of (24), it doesn't matter whom the word 'I' can pick out in a possible circumstance. Instead, we fix as the referent of a token of 'I' the actual speaker, and, and ask of this actual speaker whether she exists in some possible circumstance of evaluation. So, if the circumstances are ones in which the speaker's parents failed to reproduce; or circumstances in which the universe came to an end before the speaker was born, then in those circumstances the speaker in question obviously does not exist. Since such circumstances fail to include this speaker, it follows that, unless the speaker is a necessary existent, (24) is true in every context of utterance.

Understanding context sensitivity and its consequences for semantic analysis gives technical insight to philosophical problems. For example, they show how we can learn from statements that are metaphysically necessary. Suppose that we are indirectly watching someone moving about a room on a live video while, at the same time, we are unwittingly watching him directly moving about in front of us. After a lapse of time, we come to notice the strong similarity of the two individual's actions, and we might thereupon begin to wonder whether that guy [pointing to the man in the video] is the same as that guy [pointing directly at the same man]. After a while, let's suppose, we come to assert (26).

---

[4] Or, at least not true, since in some circumstances no agent exists, as at the center of the sun or on earth before life formed.

26. He is identical to him.

What may not yet be clear to you is that (26), indexed to the context so described, expresses a necessary truth. There are no possible circumstances under which what is said by the utterance of (26) so described might be false. Yet isn't it obvious that something informative was said with this utterance of (26), since we can wonder about truth of the utterance without incurring a charge of linguistic incompetence? This means that whatever an utterance of (26) expresses, it cannot be that we know it to be true *a priori*. So, we see how necessity and apriority can come apart. We return to such cases in §3.2 below.

But first we illustrate the complementary point that necessity and apriority can come apart the other way. Here is an example of *a priori* knowledge of a contingent truth, which has a Cartesian flavor. Anyone who understands sentence (27) *knows* that no utterance of it can be false.

27. I am here now.

Yet no utterance of this sentence or 'I exist' is *necessarily* true. Clearly, any one of us might have been somewhere other than where he currently is; or, he might have been there at a different time; or he might have been replaced there at that time by someone else, or he might not have existed at all. However, a speaker's knowledge that every utterance of (27) is true is *a priori* knowledge. Contrast (27) with sentence (28). (28) is true under exactly the same conditions as Lepore's utterance of (27).

28. Ernie Lepore is in New Brunswick at 10 a.m. on Thursday, January 27^{th}, 2000.

His knowledge or anyone else's of it is not *a priori*. Internal reflection alone can't help him decided whether he is Ernie Lepore. He might be deluded. Nor can it help me determine whether his current location is in New Brunswick or what date it currently is.[5]

---

[5] The more puzzling aspects of this sort of *a priori* knowledge might be eased if one thinks about the worlds in which these sorts of indexical statements would be false. As Descartes convincingly argued, the minimal experience required in order to consider whether these statements are true will require that you know them to be true. Those circumstances in which these sorts of statements are false are circumstances in which someone obviously cannot form the beliefs in question since he does not exist.

3.2 What sentences mean and what we understand

One of the goals of semantic analysis, as we have seen, is to show how speakers' knowledge of natural language allows them to use utterances meaningfully. Thus far, we have addressed this point only in a very preliminary way: if speakers know recursive truth conditions for their language, then they can use these rules to calculate the truth conditions for arbitrarily complex new sentences. But even if speakers can infer the new truth conditions, how exactly does this knowledge help in communication?

A preliminary, approximate answer can be found in the analogy between language and logic. We begin by idealizing the information mutually available to the interlocutors in a conversation as their common ground [Stalnaker, 1973]. The common ground settles questions whose answers are uncontroversial, in that interlocutors know the answers, know they know the answers, and so forth. Meanwhile, the common ground leaves open a set of possibilities about which there is not yet agreement: maybe there is a matter of fact that could turn out (for all the interlocutors know) to be one of various ways, or maybe the interlocutors actually do know how it turns out but do not realize the knowledge is shared — so it could be that the others know, and it could be that they don't — and so forth. Drawing on the model theory of modal logic, we might represent the possibilities in the common ground as a set of possible worlds. The model theory of modal logic likewise offers a set of possible worlds to represent the proposition associated with any formula. These are the worlds where the formula is true.

If we adopt this logical picture, we can formalize the effect that asserting a formula has on the common ground. When interlocutor A asserts a formula f, A introduces into the conversation the information that f is true. Beforehand the common ground is some set of worlds C. Afterwards the common ground also takes into account f. This formula f restricts the live possibilities by requiring the worlds in the common ground to make true a further proposition p. So the change that happens when A asserts f is that the common ground goes from C to $C \cap p$. This concise model forms the basis of a range of research characterizing the relationship between truth-conditional semantics and conversational pragmatics in formal terms; see Beaver [2001] for a recent overview.

This idealization has limits. And it's easy to come up with strange puzzles when one moves (perhaps inadvertently) beyond the limits of these idealizations. Here is such a puzzle. Why is it that a speaker would ever assert an identity statement like 'Hesperus is Phosphorus', sentence (1)? The trigger for this puzzle comes from Kripke's arguments, reviewed in §2.2, that sentence (2.3) is analytic: it is true in all possible worlds. If that is the case, consider what happens when A asserts (2.3). We update the common ground C by intersecting it with the set of all possible worlds — the proposition expressed by 'Hesperus is Phosphorus' — leaving the same set C. A has done nothing; the interlocutors' information has not changed at all! This result is absurd. What has gone wrong?

In fact, in assuming that assertions update the context with the proposition they express, we have implicitly assumed that the participants in the conversation have certain and complete knowledge of their language. For example, interlocutors can calculate that ‘Hesperus is Phosphorus’ expresses a necessarily true proposition only if they can calculate that ‘Hesperus’ names Venus and ‘Phosphorus’ names Venus. Of course, under such circumstances, they don’t learn anything from the sentence. But it is also easy to see how this assumption could go unnoticed in extrapolating from logic to language. When we do logic, we are typically interested in inferential relationships that hold no matter what our terms refer to. We assume that the reference of our terms — precisely what matters in the ‘Hesperus is Phosphorus’ case — is not at issue.

Let us consider how to formalize uses of sentences in more realistic situations. As we do so, we have to be very careful to respect the intuitions of Kripke and Putnam’s thought experiments [Stalnaker 1978].

Suppose interlocutor B does not know that Hesperus is Phosphorus. Recall from our discussion of (2.3) in §2.2 that what that really means is that B cannot distinguish between two possible situations. In the first, there is only one heavenly body out there, and B and B’s community speak a language English$_1$ where both *Hesperus* and *Phosphorus* are rigid designators for that body. In the second, there are two distinct heavenly bodies, and B speaks a language English$_2$ where *Hesperus* is a rigid designator for one of them and *Phosphorus* is a rigid designator for the other. Since these possibilities are open for B, they must both also be represented in the common ground.

Now, we need a correspondingly expressive notion of assertion. When interlocutor A says something, A is committed that it is true according to the standards for reference that prevail in the community. Any assertion that A makes should turn out to be true in the language A speaks. What we have just seen is that any point of evaluation w in the common ground could potentially have its own language English$_w$ with relevant differences from English as spoken in the real world. Adapting Stalnaker’s [1978] terminology,[6] we can associate any utterance u with a *diagonal proposition*: this proposition is true at a point w if the proposition that u expresses in English$_w$ is true in w. In the case of *Hesperus is Phosphorus*, the effect of A’s assertion is to intersect the common ground with this diagonal proposition. Concretely, we retain in the common ground worlds of the first kind, where English$_1$ is spoken, Hesperus and Phosphorus are necessarily the same and A's assertion is necessarily true. However, we discard from the common ground worlds of the second kind, where English$_2$ is spoken, Hesperus and Phosphorus are necessarily different and A's assertion is necessarily false.

---

[6] Stalnaker’s presentation of the idea is rather more complicated. Stalnaker’s idea is to use possible worlds as representations of the context of an utterance, and to treat the reference of proper names as a kind of context-sensitivity. We think it’s better to distinguish between interlocutors’ epistemic context, which describes the information they share, and their linguistic context, which describes the state of their interaction. For example, the interaction could have a well-defined state that only one of the participants knows about; this would be part of the linguistic context but not the epistemic context. Analogously, we think it’s better to distinguish between interlocutors’ uncertainty *about* their language and the context-sensitivity of terms *within* their language. Stalnaker concludes that it is a kind of Gricean repair when an utterance conveys the diagonal proposition in his sense of factoring out all kinds of context-dependency; but we see no reason why an utterance should not always convey the diagonal proposition in our sense of factoring out dependence on the language spoken.

There is substantially more to be said about the relationship between utterance meaning and the information that interlocutors convey. Speakers are rational, so they can communicate indirectly through the choices they make about what information to give and how to present it [Grice, 1975]. And language use is collaborative, so speakers can often count on hearers to accommodate information they implicitly presuppose [Lewis, 1979]; [Thomason, 1990] and to recognize projects they indirectly suggest [Searle, 1975]. The element of strategy continues to inform fine-grained accounts of linguistic phenomena at the boundary between semantics and pragmatics — for everything from the resolution of pronouns [Beaver, 2004] to the use of vague language [Kyburg and Morreau, 2000]; [Barker, 2002]. Throughout this research, as in the example from Stalnaker presented above, the need to consider the knowledge and choices of actual speakers remains a central theoretical construct that distinguishes semantic analysis from logic.

4 DOING SEMANTIC ANALYSIS IN LOGIC

Logic offers more than an analogy for doing semantics for natural language. Logic is a tool that makes semantic analysis easier to do, to present and to understand. It allows us to be precise about the ontology that our semantics presupposes — what knowledge of different kinds of entities and relationships best explains our ability to use language. It allows us to hook up that ontology precisely with the syntax. And it allows us to use the mathematics of logical consequence to describe the semantic relationships among sentences. This mathematics shows how our judgments might be grounded in computations over representations of knowledge of meaning. In §4.1, we describe one of the most influential theories of the sort, Davidson's [1967a; 2002] use of events to specify the logical form of verbs of change. Such philosophical analyses are important to semantics, but, as we emphasize in §4.2, these analyses are most successful when they are most banal. We are generally skeptical that semantic analyses that explain how word meanings compose to yield sentence meanings will lead to substantive insights about the internal structure of word meanings themselves. Thus, the project of semantic analysis needs to be clearly distinguished from the broader project of using logical formalisms to model mental representations and cognitive processes in general.

4.1 Verbs of change and their modifiers[7]

Verbs of change are verbs that correspond to events, processes or actions which occur, go on, or take place. For example, the verbs 'to leave' and 'to kiss' are verbs of change. Such verbs can be modified by prepositional phrases, as in 'John left *for* the store' or 'Frank kissed Mary *in* Tompkins Square'. This sort of modification seems productive, since there is no obvious limit to how many modifying prepositional phrases we can modify a verb of change with, as the sequence of sentences (29) through (33) suggests.

29. Mary kissed John.
30. Mary kissed John in Tompkins Square.
31. Mary kissed John in Tompkins Square after midnight.
32. Mary kissed John in Tompkins Square after midnight behind his left ear.
33. Mary kissed John in Tompkins Square after midnight behind his left ear on August 24, 1999.

If this process of expansion can be continued indefinitely, then English can express an open-ended set of propositions just by virtue of adding prepositional phrases to a sentence that modify the verb.

Interpreting (29)–(33) thus requires reaching into our bag of learned expressions in order to compose their meanings into meaningful complexes. One promising proposal for how to do this is provided by what we shall call the *event analysis.* According to this proposal, (29) and (30) are best understood to be synonymous with (34) and (35).

34. *There is an event that was a kissing by* Mary *of* John.
35. *There exists an event that was a kissing by* Mary *of* John *and that was in Tompkins Square.*

That is, events figure explicitly in the rules and computations English speakers use to formulate and understand sentences like (29) and (30).

This is a substantive discovery about the meaning of English. Given their surface form, you might have thought that (29) involved a two-place predicate 'kissed' relating Mary and John, and you might have thought that (30) involved a three-place predicate 'kissed in' relating Mary, John and the park. However, the analyses in (34) and (35) urge us *not* to give 'kissed' these surface semantics. Instead, the analyses use the same underlying three-place predicate, '_ was a kissing by _ of _', both for (29) and for (30). This predicate makes the event of kissing explicit. The prepositional phrase 'in Tompkins Square' on this account contributes the predicate '_ was *in* Tompkins Square'; the prepositional phrase 'after midnight' contributes the predicate 'it was *after* midnight' and so on. The event approach solves the problem of productivity by introducing events about which indefinitely many things can be said.

---

[7] This section is largely a summary of Davidson [1967a; 2002].

Another distinctive feature of the event analysis is that it describes the semantics of action sentences in terms of an existential quantifier phrase — 'there is an event' in (34) and (35). This quantifier takes scope over the entire action description, containing the verb and all its modifiers. This quantificational structure correctly predicts that the verb and all its modifiers describe one and the same event, while no constituent in the sentence explicitly identifies or names that event. As we shall see, the event analysis also draws on this added structure to explain what inferences the resulting sentences support, and how they function in connected discourse.

The formal implementation of this event analysis provides a good illustration of some of the contributions of logic to the practice of semantic analysis. To start, suppose we use the formalism of higher-order logic as a meta-language for doing natural language semantics: then we can start to capture the event analysis of (34) and (35) in the formalism by introducing the symbolism *kissed(E,X,Y)* to indicate that event E was a kissing by X of Y. This mnemonic is useful not only because it is short, but also because it highlights one of the claims of the theory, namely that whenever the verb 'kissed' appears in English it expresses a relationship that underlyingly characterizes an event. The notation suggests, in a way that paraphrases like (34) and (35) might not, that the events need to be made explicit to give any systematic semantics for sentences with the verb 'kissed'. Letting M be Mary, J be John and T be Tompkins Square, we now have these symbolic representations for the truth conditions of (29) and (30):

36. $\exists\, E\ (kissed(E,M,J))$.

37. $\exists\, E\ (kissed(E,M,J) \wedge in(E,T))$.

By symbolizing the contribution of 'in Tompkins Square' using the logical term *in(E,T)* we are able to concisely indicate that the kissing event is what is being described and that what is being said is that the event was in the park. Again, the notation emphasizes the claimed analysis, namely that 'in' always expresses a relationship between a thing (in this case an event) and a spatial landmark that locates it.

We can develop the formalism further to show how the event analysis explains why we can keep adding prepositional phrases to modify verbs of change without a loss of grammaticality; it does so by treating such additions through the mechanism of predicate conjunction. Predicate conjunction is a ubiquitous semantic operation [Partee and Rooth, 1983; Chung and Ladusaw, 2003]; you can clearly see the need for it in examples such as (38)–(40).

38. Felix is a round table.

39. Felix is a brown round table.

40. Felix is a wooden brown round table.

We could continue to add adjectives without a loss of grammaticality indefinitely. It is standard to treat this sort of modification as conjunctive. (38) is true just in case Felix is both round *and* a table. So even though grammar legislates that 'round' in (38) is an adjective that modifies 'table', since (41) paraphrases (38),

41. Felix is a table *and* Felix is round.

where 'table' and 'round' overtly describe Felix in (41), they must do so covertly as well in (38). If you understand 'round' and 'table', all you need to learn to understand 'round table' is that it applies to an object just in case both the simpler component expressions apply to it as well. The general rule from which this specific case derives is (PA).

> (PA) A complex common noun devised by prefacing a common noun (itself possibly complex) with an adjective applies to an object just in case *both* the adjective *and* the common noun apply to that object.

So (PA) applies to (40) three times, and tells us that overall (40) is true just in case 'wooden' and 'brown' and 'round' and 'table' all apply to Felix. The same kind of mechanism applies in composing (36) and (37). We first build a complex predicate characterizing an event — that it was a kissing by Mary of John, that it was in Tompkins Square — and finally derive the sentential condition that there is some event that satisfies this complex predicate.

Using a logical vocabulary also helps us demystify these compositional derivations. A complex conjunctive statement — 'John is tall and Mary is happy' — is true if and only if its simpler conjuncts — 'John is tall', 'Mary is happy'— are true. Letting *tall(J)* and *happy(M)* represent the content of the individual conjuncts allows us to represent the content of the conjunctive statement by *tall(J)* $\wedge$ *happy(M)*. Formalism in our meta-language here highlights our commitment to ordinary logical conjunction. For example, it allows us to carry out inferences on our representation using the deductive operation of conjunction reduction. From the fact that a sentence of the form $\ulcorner P \wedge Q \urcorner$ is true, we can infer that 'P' is true and we can infer that 'Q' is true.

Analogously, a complex conjunctive predicate — 'is tall and happy'— is true of an object if and only if its conjuncts — 'is tall', 'is happy'— are as well. Now, higher-order logic lets name predicates using the formalism of λ-abstraction: λ X. *tall(X)* names the predicate that is true of an object X if X is tall. In general $\lambda X.C$ names the function which gets its result on argument A by evaluating C on the assumption that $X = A$. This equivalence is embodied in the λ-calculus equivalence $(\lambda X.C)(A) \equiv C[X \leftarrow A]$. For example, ($\lambda$ X. *tall(X))(f)* $\equiv$ *tall(f).*

With this notation, we can represent the conjunctive predicate with the λ-term λ X. *tall(X)* $\wedge$ *happy(X).* In general, we can represent the conjunction of the predicates F and G by the λ-term $\lambda X.F(X) \wedge G(X)$. Thus this notation allows us to rewrite (PA) compactly in a way that shows its close similarity to the interpretive rules illustrated in (5)–(7). We give this new rule as (PAL).

> (PAL) If N is a (complex) common noun that refers to the predicate G and A is an adjective that refers to the predicate F, then $\ulcorner AN \urcorner$ is a complex common noun that refers to the predicate conjunction $\lambda X.F(X) \wedge G(X)$.

This formalism again allows us to highlight our commitment to ordinary logical conjunction. Ultimately the same meaning of ‘and’ is at play whether we are conjoining sentences or predicates. For example, the expected inferences follow: If a complex predicate $\ulcorner \lambda X.P \wedge Q \urcorner$ is true of some object A, in other words $\ulcorner (\lambda X.P \wedge Q)(A) \urcorner$ is true, then $\ulcorner P[X \leftarrow A] \wedge Q[X \leftarrow A] \urcorner$ follows (by λ-calculus equivalence) and so both $\ulcorner P[X \leftarrow A] \urcorner$ and $\ulcorner Q[X \leftarrow A] \urcorner$ follow (by conjunction reduction) and thus both $\ulcorner \lambda X.P \urcorner$ is true of A and $\ulcorner \lambda X.Q \urcorner$ is true of A (by λ-calculus equivalence, again).

Now, according to the event analysis, the conjunctive predicate ‘was a kissing by John of Mary *and* was in Tompkins Square’ is true of an event just in case its simpler component predicates ‘was a kissing by John of Mary’ and ‘was in Tompkins Square’ are both true of that event as well. Formally, we represent the conjoined predicate as λ *E. kissed(E,M,J)* $\wedge$ *in(E,T)* and derive it by predicate conjunction from λ *E. kissed(E,M,J)* and λ *E. in(E,T).*

Another key ingredient of the proposal is the existential quantification over events (in both (34) and (35)) in their main verb and (in (35)) in its preposition. The latter two sentences tell us to treat the former two as existentially quantifying over an event. An initial response to this suggestion might be doubt, since (29) and (30) do not seem to *assert* the existence of an event; i.e., they don't seem to harbor any existential quantifier. But in fact, the patterns of inference of action sentences are exactly what you would expect from the existential analysis. For example, (42) does not entail (43), but (43) does entail (42):

42. Mary kissed John in Tompkins Square and Mary kissed John on Wednesday.
43. Mary kissed John in Tompkins Square on Wednesday.

This is just what you would expect if the meaning of (42) is represented with two existential quantifiers as in (44) and the meaning of (43) is represented with one as in (45).

44. $\exists$ *E (kissed(E,M,J)* $\wedge$ *in(E,T))* $\wedge \exists$ *E (kissed(E,M,J)* $\wedge$ *on(E,W))*
45. $\exists$ *E (kissed(E,M,J)* $\wedge$ *in(E,T* $\wedge$ *on(E,W))).*

Patterns of discourse connectivity are also exactly what you would expect from the existential analysis. The familiar pronoun ‘it’ in (46) seems to be about something, presumably an action.

46. John buttered his toast and he did *it* after midnight.

It is reasonable to ask whether the pronoun ‘it’ in (46) anaphoric or deictic? Since (46) is intelligible without a concomitant pointing gesture or any other sort of demonstration, ‘it’ must be anaphoric. But what is ‘it’-s anaphor? The other noun phrases in (46) are ‘John’, ‘his toast’, ‘he’ and ‘midnight’, but these are not actions, and so ‘it’ neither refers back to nor is bound by any of them. Since nothing else is explicitly referenced in (46), we must understand its pronoun ‘it’ to be anaphoric. But what can its anaphor be? The event approach provides a simple answer: it interprets the verb ‘butter’ in (46) as having a place quantifying over events. (46) is understood as (47).

47. Some event was a buttering of the toast by John and *it* (i.e., the event) occurred after midnight.

(47) explains why ’it’ in (46) is intelligible. It is an anaphoric pronoun whose anaphor is ‘some event’. The event approach posits an *implicit* position in verbs *and* the behavior of ‘it’ in (46) provides the right evidence for this posited place. The support is unexpected and intuitively pleasing. Data of the sort provided by (46) reminds us that part of what we learn in learning the meaning of a verb of change (indeed, any verb) is its adicity, and what sorts of entities the expressions which fill these places are true of. Verbs of change, it turns out, have an event place.

As explanatorily rich as the event account is it still faces problems. As developed thus far it yields unwanted inferences if applied generally. To see this, consider the distinction between the modifiers ‘quickly’ and ‘allegedly’. The adverb of manner ‘quickly’ in (48) is modifying the verb ‘drove’.

48. Frank drove quickly.

The manner in which the adverb modifies the verb in (48) renders it impossible for (48) to be true and (49) false.

49. Frank drove.

This sort of inference from a sentence with adverbial modification to one without it holds in general. The inference is not only valid but valid *in virtue of its form*, much like inferential relations among (29) through (33).

By contrast, if we symbolize (50) along the lines recommended by the event approach, it winds up logically implying (51).

50. Clinton *allegedly* lied under oath.

51. Clinton lied under oath.

The interpretation of (50), according to the event approach, is (52).

52. There is an event e such that e was a lying under oath by Clinton *and* e was alleged.

But alleging that something is all by itself doesn't make it so. This may be a problem for the event approach, for it focuses on a somewhat special class of modifiers, 'quietly' as opposed to 'allegedly', 'in Tompkins Square' as opposed to 'in a dream'.

Should we ditch the event approach on this basis and seek a more general approach? Before sacrificing it, a closer look at so-called 'problem' modifiers, for example, 'allegedly', suggests a non-*ad hoc* way of distinguishing problem cases from non-problem ones. We can move from (53) to (54) without a change in meaning.

53. Oedipus allegedly married his mother.

54. It has been alleged that Oedipus married his mother.[8]

Since (53) and (54) are synonymous, we can infer that 'allegedly' is not modifying the verb 'married' in (53), but rather, as in (54), 'it has been alleged that' is predicated of a statement. This makes sense, since it is statements that are alleged. And, so this transformation works, because a modifier like 'allegedly' is not an ad*verb* modifying a verb, but instead it modifies whole sentences. To take another example, when you say 'John *clearly* studied for this exam' you are not attributing clarity to his extended event of studying for this exam. Rather, you are asserting '*It is clear that* John studied for this exam'. What is clear is the statement that John studied for this exam. In this regard, adverbs like 'allegedly' and 'clearly' differ from 'quickly' and 'quietly'. Transforming (48) into (55) does not preserve grammaticality.

55. *It has been quick that Frank drove.

If correct, we may have a principled distinction between adverbs the event approach accommodates and those it cannot. What we do not know yet is, if not as conjoined predicates attributed of events, what sort of contribution do adverbs like 'allegedly' and 'clearly' make to sentences in which they occur. Whatever contribution it is, it will be distinct from the sort of contribution an adverb like 'quickly' makes, for suppose that Frank is identical to your uncle, then if Frank drives quickly, so too does your uncle. But even though Oedipus' mother was Jocasta, no one alleged that Oedipus married Jocasta. If this is right, then adverbs like 'allegedly', 'clearly', 'intentionally', 'on purpose' cannot be treated as predicates of events or actions at all. What, then, is their contribution to meaning? We cannot hope to settle all these issues here.[9]

---

[8] 'Supposedly John left' ('It is supposed that John left'), 'John killed Bill in Fred's dream' ('Fred dreamt that John killed Fred').

[9] For further discussion see [Parsons, 1990].

4.2 Limits to analysis

It is unchallengeable that the event approach is a paradigm of a substantial analysis about a systematic compositional meaning theory for natural language. Even though these gains are nothing to sneer at, it is an unsettling exchange we have transacted — one that highlights the holistic, scientific character of our investigation into meaning. In effect, we have bargained that if the event approach explains various logical relations and accounts for linguistic productivity, we will ignore our intuitions about the sentences in question. We will ignore our intuition that the seemingly two-place verb ‘_ kissed ’ in (29) through (33) is elliptic for a three-place verb ‘_ was a kissing by _ of ’. Granted, this consequence is not as bad as the discredited proposal that left the adicity of each verb of change as an open-ended affair. But even though on the event analysis there is a fixed number of places for each verb of change, the number that the approach posits is not what we pre-theoretically think it is. Also, whatever contribution prepositions make to the meaning of sentences in which they occur, we do not pre-theoretically think that it's the same sort of contribution that predicates make. Yet, on this proposal, prepositions in (29) through (33) turn out to be two-place predicates ('in' in (30) through (33) is elliptic for the two-place predicate ‘ is in ’; and so forth). And, lastly, prior to encountering the event approach, it would never have occurred to an English speaker to conclude that (29) through (33) harbor hidden existential quantifiers. So in order to endorse the event approach we had to defend these three surprising and unintuitive consequences. In the end, we can muster a wide variety of evidence in support of these consequences of the event analysis — despite its surprises, the event analysis offers the best account available.

Many proposed analyses in semantics are not such clear wins. For example, analyses that purport to give logical equivalents for lexical meanings are particularly problematic. Paradigmatic examples of such analyses include the hypothesis that (56) is synonymous with or elliptical for (57), that (58) is synonymous with or elliptical for (59), and that (60) is synonymous with or elliptical for (61).

56. Chris used stain to darken the wood.

57. Chris used stain to cause the wood to become darker.

58. Kim boiled the water.

59. Kim caused the water to boil.

60. The tiger killed the elephant.

61. The tiger caused the elephant to die.

At first, these analyses might seem to offer comparable advantages to the event analysis. For one thing, these analyses immediately account for some important inferential relationships among sentences. It's clear, for example, that (56) definitely entails (57), that (58) definitely entails (59), and that (60) definitely entails (61).

For another thing, these analyses, like the event analysis, explain apparently productive processes in language. In English, the '-en' suffix that appears on 'darken' works more generally to turn an adjective X into a verb meaning, roughly, 'make more X'. It applies to lots of adjectives (though only to short Anglo-Saxon roots ending in consonants of low sonority). Similarly, a wide variety of verbs describing processes things undergo, like 'boil', can also be used in a different sense meaning, roughly, to cause something to undergo the process.

On balance, however, we don't think the preponderance of the evidence favors these analyses. The productivity of these constructions is limited. There is a qualitative difference between the infinite productivity of open-ended modification and the vast but still finite productivity of a lexical derivation that applies to any root word in the language. Speakers have to learn each root form — root forms are arbitrary pairings of sound and meaning, after all — so it's certainly possible that speakers could learn something about each derived form, too. In fact, it's common for complex words to have *motivated* meanings that are related to the meanings of the parts but not predictable from them. The noun 'wheelchair' is a particularly clear case: the word contains the words 'wheel' and 'chair', and, as you might expect, all wheelchairs are chairs on wheels. Yet 'wheelchair' doesn't just mean chair on wheels: it refers to a specific common kind of chair with wheels. It is a reasonable hypothesis that the verb meanings for 'darken' in (56) and 'boil' in (58) are motivated in this sense: they contain a morpheme that expresses causality, but have a more specific meaning that reflects this causality in this semi-compositional way. You can cause something to be darker without darkening it. Turning down the dimmer switch on a light bulb makes the bulb darker but we don't call that 'darkening the bulb' in English. Lowering the pressure in a container of water can make the water boil but isn't boiling the water. And it is also possible to cause something to die without killing it; Romeo's suicide causes Juliet to die, but she kills herself.

These mismatches are reminders that providing analyses in terms of covert logical structure is generally a counterintuitive business. Intuitively, killing is killing and dying is dying — different kinds of things, though certainly related. The same goes for the boiling that you do and the boiling the water does, or the darkening that you do and the dark wood that results — these are different kinds of things, though certainly related. In many cases, such relationships can be captured in lots more transparent ways than by analyses that assume that language contains complex hidden structure.

5 OPEN PROBLEMS AND FURTHER READING

In this chapter, we have suggested that philosophy has a place for truths about meaning, and have explored logic as a tool for developing these truths. Throughout, our discussion we have emphasized both the strong points of this program and its limitations.

On our view, truths about meaning can play an important role in the philosophical description of language. We can use our knowledge of meaning to understand sentences and to communicate with one another. However, truths about meaning may have rather less import for the rest of philosophy — for questions in epistemology or metaphysics, for example — because facts about meaning won't come with *a priori* certainty. Of course, the particular ways in which we discover these facts may help us clarify our philosophical arguments, and while such possibilities are beyond the scope of this chapter, ongoing debates certainly explore them. A case in point is the contextualist semantics for *know* where uses of the verb attribute knowledge not absolutely but only with respect to prevailing epistemic standards [Lewis, 1996].

On our view, logic provides an important collection of tools that we can take as models for the semantic analysis of natural language. The logical treatment of sentential operators, functions and arguments, variables, quantifiers and scope, tense and modality have all inspired research in natural language semantics. There are many good introductions to the overlap between logic and language, particularly [Gamut, 1991], [Kamp and Reyle, 1993], and [Larson and Segal, 1995]. There is even a *Handbook of Logic and Language* [van Benthem and ter Meulen, 1997]. Technical sophistication comes in particularly handy when the theorist aims to establish a direct correspondence between fine-grained linguistic structure and a compositional theory of truth; so logic is a mainstay of formal linguistic semantics. At the same time, however, we have seen how pragmatic dimensions of language use — context dependence and context change — may motivate extensions to or departures from mathematical logic in semantic analysis. *The Handbook of Pragmatics* [Horn and Ward, 2004] provides an entrée into the diverse phenomena that characterize language in use.

Finally, however, we emphasize that semantic analysis is an empirical discipline that aims to reconcile our intuitions about meaning with our abilities to use language, draw inferences, evaluate arguments and communicate ideas. Philosophy of language provides a framework for this science and logic provides a tool — in many respects, this is where the action finally starts. Fundamental issues remain controversial: what kinds of entities are required for semantic theory? How do these entities get hooked up with our words and utterances? What constraints must a semantic theory for natural language respect? And what is the role of these constraints and our own experience in allowing us to acquire our knowledge as speakers of our native language? We have highlighted the opportunities and the challenges of these questions in semantic analysis here, with the example of action verbs. But they are as open-ended as language itself.

BIBLIOGRAPHY

[Areces *et al.*, 2001] C. Areces, P. Blackburn, and M. Marx. Hybrid logic: Characterization, interpolation and complexity. *Journal of Symbolic Logic*, 66(3): 977–1010, 2001.

[Barker, 2002] C. Barker. The dynamics of vagueness. *Linguistics and Philosophy*, 25(1): 1–36, 2002.

[Beaver, 2001] D. Beaver. *Presupposition and Assertion in Dynamic Semantics.* CSLI Publications, Stanford CA, 2001.

[Beaver, 2004] D. Beaver. The optimization of discourse anaphora. *Linguistics and Philosophy*, 27(1): 3-56, 2004.

[van Benthem and Meulen, 1997] J.F.A.K. van Benthem and G.B.A. ter Meulen (eds.). *The Handbook of Logic and Language.* MIT Press, Cambridge, Mass, and North-Holland, Amsterdam, The Netherlands, 1997.

[Chung and Ladusaw, 2003] S. Chung and W. Ladusaw. *Restriction and Saturation.* MIT Press, Cambridge Mass, 2003.

[Davidson, 1967a] D. Davidson. The logical form of action sentences. In N. Rescher (ed.), *The Logic of Decision and Action*, 81-95. University of Pittsburgh Press, Pittsburgh, 1967. Reprinted with discussion in Davidson [2002], 105-148.

[Davidson, 1967b] D. Davidson. Truth and meaning. *Synthese*, 17: 304-323, 1967. Reprinted in Davidson [2001], 16-36.

[Davidson, 2001] D. Davidson. *Inquiries into Truth and Interpretation.* Second Edition, Clarendon Press, Oxford, 2001.

[Davidson, 2002] D. Davidson. *Essays on Actions and Event, Second Edition.* Clarendon Press, Oxford, 2002.

[Frege, 1879] G. Frege. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Halle: L. Nebert, 1879. Translated as *Begriffsschrift, a Formula Language, Modeled upon that of Arithmetic, for Pure Thought.* In *From Frege to Gödel*, edited by J. van Heijenoort. Cambridge, MA: Harvard University Press, 1967. Also as *Conceptual Notation and Related Articles.* Edited and translated by T.W. Bynum. London: Oxford University Press, 1972.

[Fitting and Mendelsohn, 1998] M. Fitting and R.L. Mendelsohn. *First-order Modal Logic.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[Fodor, 1975] J.A. Fodor. *The Language of Thought.* Crowell, New York, 1975.

[Fodor, 1987] J.A. Fodor. *Psychosemantics.* MIT Press, Cambridge Mass, 1987.

[Gamut, 1991] L.T.F. Gamut. *Logic, Language, and Meaning*, Vol. 1: *Introduction to Logic*, and Vol. 2: *Intensional Logic and Logical Grammar.* University of Chicago Press, 1991.

[Grice, 1975] H.P. Grice. Logic and conversation. In P. Cole and J. Morgan (eds.), *Syntax and Semantics III: Speech Acts*, 41–58. Academic Press, New York, 1975. Reprinted in Grice [1989], 22-40.

[Grice, 1989] H.P. Grice. *Studies in the Way of Words.* Harvard University Press, Cambridge, Mass, 1989.

[Groenendijk and Stokhof, 1990] J. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14: 39–100, 1990.

[Hinrichs, 1986] E. Hinrichs. Temporal anaphora in discourses of English. *Linguistics and Philosophy*, 9(1): 63–82, 1986.

[Horn and Ward, 2004] L.R. Horn and G. Ward (eds.), *Handbook of Pragmatics.* Basil Blackwell: Oxford. 2004.

[Kamp and Reyle, 1993] H. Kamp and U. Reyle. *From Discourse to Logic.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.

[Kaplan, 1989] D. Kaplan. Demonstratives. In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, 481-566. Oxford University Press, Oxford, 1989.

[Kripke, 1972] S.A. Kripke. Naming and necessity. In D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, 253–355. Reidel, 1972. Reprinted as Kripke [1980].

[Kripke, 1980] S.A. Kripke. *Naming and Necessity.* Harvard University Press, Cambridge Mass, 1980.

[Kyburg and Morreau, 2000] A. Kyburg and M. Morreau. Fitting words: Vague words in context. *Linguistics and Philosophy*, 23(6): 577–597, 2000.

[Larson and Segal, 1995] R. Larson and G. Segal. *Knowledge of Meaning: An Introduction to Semantic Theory.* MIT Press, 1995.

[Lewis, 1979] D. Lewis. Scorekeeping in a language game. In R. Bäuerle, U. Egli, and A. von Stechow (eds.), *Semantics from Different Points of View*, 172–187. Springer Verlag, Berlin, 1979.

[Lewis, 1996] D. Lewis. 'Elusive Knowledge'. *Australasian Journal of Philosophy*, 74. Reprinted in *Papers in Metaphysics and Epistemology*, 549–67, 1996.

[Montague, 1970] R. Montague. English as a Formal Language. In B. Visentini *et al.* (eds.), *Linguaggi nella società e nella tecnica.* Milan: Edizioni di Comunità. 189-224, 1970. Reprinted in Montague [1974], 188-221.

[Montague, 1973] R. Montague. The proper treatment of quantification in ordinary English. In K.J.J. Hintikka, J.M.E. Moravcsik, & P. Suppes (eds.) *Approaches to Natural Language.* Dordrecht: Reidel. 221-242, 1973. Reprinted in Montague [1974], 247-270.

[Montague, 1974] R. Montague. *Formal Philosophy. Selected Papers of Richard Montague.* Edited and with an introduction by R.H. Thomason. New Haven/London: Yale University Press, 1974.

[Muskens, 1996] R. Muskens. Combining Montague semantics and discourse representation theory. *Linguistics and Philosophy*, 19(2): 143–186, 1996.

[Parsons, 1990] T. Parsons. *Events in the Semantics of English: A Study of Subatomic Semantics.* MIT Press, Cambridge, 1990.

[Partee, 1973] B.H. Partee. Some structural analogies between tenses and pronouns in English. *Journal of Philosophy*, 70: 601–609, 1973.

[Partee, 1984] B.H. Partee. Nominal and temporal anaphora. *Linguistics and Philosophy*, 7(3):243–286, 1984.

[Partee and Rooth, 1983] B.H. Partee and M. Rooth. Generalized conjunction and type ambiguity. In R. Bauerle, C. Schwarze and A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*, Walter de Gruyter, Berlin, 361-383, 1983.

[Pietroski, 2006] P. Pietroski. Logical form and LF. In E. Lepore and B. Smith (eds), *Oxford Handbook to Philosophy of Language.* , Oxford: Oxford University Press, 2006.

[Putnam, 1976] H. Putnam. The Meaning of 'Meaning'. In *Philosophical Papers II: Mind, Language, and Reality*, Cambridge: Cambridge University Press, 215-271, 1975.

[Quine, 1953] W.V.O. Quine. Two Dogmas of Empiricism. In *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press, 20-46, 1953.

[Quine, 1960] W.V.O. Quine. *Word and Object.* Cambridge, Mass.: The Technology Press of M.I.T; New York, 1960.

[Robert, 1989] C. Roberts. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy*, 12(6): 683–721, 1989.

[Searle, 1975] J.R. Searle. Indirect speech acts. In P. Cole and J. Morgan (eds.), *Syntax and Semantics III: Speech Acts.* Academic Press, New York, 59–82, 1975.

[Stalnaker, 1973] R. Stalnaker. Presuppositions. *Journal of Philosophical Logic*, 2(4): 447–457, 1973.

[Stalnaker, 1978] R. Stalnaker. Assertion. In P. Cole (ed.), *Syntax and Semantics* 9: 315-332. Academic Press, New York, 1978. Reprinted in Stalnaker [1999], 78-95.

[Stalnaker, 1999] R. Stalnaker. *Context and Content.* Oxford University Press, Oxford, 1999.

[Tarski, 1935] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen, *Studia Philosophica*, I: 261-405, 1935. Translated into English and published under the title 'The Concept of Truth in Formalized Languages' in Tarski [1956].

[Tarski, 1956] A. Tarski. *Logic, Semantics and Metamathematics.* Oxford University Press, 1956.

[Thomason, 1990] R.H. Thomason. Accommodation, meaning and implicature. In P.R. Cohen, J. Morgan, and M.E. Pollack (eds.), *Intentions in Communication*, 325–363. MIT Press, Cambridge, Mass, 1990.

[Webber, 1988] B.L. Webber. Tense as discourse anaphor. *Computational Linguistics*, 14(2): 61–73, 1988.

JUSTIFICATORY IRRELEVANCE OF FORMAL SEMANTICS

Charles F. Kielkopf

1 BACKGROUND AND OVERVIEW

The stimulus for this essay was an invitation from Professor Dale Jacquette to submit a paper on validity and entailment for this volume on “Philosophy of Logic.” I welcomed the invitation as an opportunity to strengthen a case for system E of Entailment as the correct formal sentential logic. In 1977, I had published a survey of several systems of formal sentential logics along with semantics, or notions of validity, associated with the systems.[1] In that book I defended system *E* of entailment as the correct formal sentential logic. I tried to show that a natural language reading of the formulas of *E* best preserved widely held intuitions, or prejudices, about what ought to be accepted and rejected as legitimate deductive inference. Despite following the fashion of the 1960s and 70s by comparing many competing systems of logic, I did not, and still do not, accept that there can be a plurality of correct logics. Logic consists of our practices for settling disputes about correct inferences. If reasonable people can argue about the correctness of the practices the practices are not logic even if they are useful and philosophically supported formalized practices for rigorously drawing inferences from premises. In this essay I focus exclusively on sentential logic.

I was uneasy with my case for system *E* because it appealed to the implicit semantics we use in understanding the sentences of our natural languages. Such appeals seem so similar to appeals to prejudices. Still, I believed that a case that a particular use of logical terms such as ‘or,’ ‘and,’ ‘not,’ ‘if_ then_,’ etc., conforms to how these terms ought to be used would require an appeal to semantics of some type. I had an inchoate hope that somehow recourse to formal semantics with a formal notion of validity would make the case rigorous. In a first response to Professor Jacquette’s invitation I made a vain effort to realize that hope. I tried to interpret the axioms for *E* model structures as true claims about entailment and the associated notion of validity as expressing what we really mean by ’validity.’ Unfortunately, acceptance of these notions of validity as the sole proper notion of validity becomes as much a matter of dispute, and prejudice, as acceptance of the axioms of the system read simply as rules for the sole proper system of inferences. So here I submit a negative conclusion that formal semantics are not useful

---

[1] [Kielkopf, 1977].

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

for deciding which system of formal logic is the correct system; in effect, formal semantics are part of the system to be justified. This negative conclusion should have been immediately obvious to me long ago. Fortunately, the formal semantics can be used, within limits, to make conclusive metalinguistic determination about what is proved or, on occasion what is not provable, in a system such as E of entailment, I of intuitionism, C of classical logic, etc. So here, I propose that we change, at least for philosophical discussions about the right or correct logic, the notion of the fundamental logic for formal reasoning. Fundamental, or genuine logic, is constituted by those metalinguistic practices with which we determine the correctness of reasoning within a system by reference to the standards proposed for the system. Ideally rational people will not disagree about the application of these practices to evaluate formal reasoning and definitely these practices must not require justification by appeal to semantic intuitions about the logical constants. Exhibition of some of these practices support labeling them with the vague term "strict finitistic."

On my proposal, truths of fundamental sentential logic are claims about what is formally provable in various formal systems for sentential logic. If a proof has been presented in a system S of Ψ from Γ, i.e., $\Gamma \vdash_s \Psi$ and the sequence of formulas comprising the proof is surveyable, typical people who know about formula proofs can determine whether or not $\Gamma \vdash_s \Psi$ is true. If the system has a decision procedure, then if the application of the decision procedure is not too complicated, we can make a metalinguistic determination that certain claims of the form $\Gamma \vdash_s \Psi$ cannot be true. For some systems without a decision procedure, such as E, a surveyable matrix can occasionally be used to show that an $\Gamma \vdash_E \Psi$ cannot be true. We shall see how E model structure semantics are heuristics for constructing reasonably small matrices. My methodology is Wittgensteinian by virtue of following the maxim that philosophy is assembling reminders for a particular purpose. Work with a few procedures in formal semantics remind us that these procedures are irrelevant to justifying as the correct logic the formal systems associated with them. Secondly, I exhibit what we can accomplish with these semantics to justify truth claims about formal provability. These second reminders show that we can frequently say with mathematical precisions what is and is not provable in various systems.

If we accept the proposal that "fundamental" logic is constituted by the strict finitistic claims about the syntax and semantics and classify strict finitistic claims as mathematical, then logic becomes mathematics. So instead of a logicistic thesis that mathematics is logic, this essay proposes what could be called a mathematist thesis that logic is mathematics.

2 THE PROBLEM OF JUSTIFYING A FORMAL LOGIC

I follow an early 20^{th} century practice guided by Hilbert, Carnap[2] *et al.* and use an idealized formal presentation of science to locate the problem of justifying a formal logic. I assume familiarity with the syntactical preliminaries for setting up a formal system and characterize the idealization by citing the types of axioms in a very rich formal system.

The idealization:

i. Axioms and rules of proof for sentential logic

ii. Axioms and rules of proof for predicate logic without =

iii. Axioms for =

iv. Axioms and rules for modalities, if used

v. Axioms for set-theory and other mathematics if not obtainable from set theory.

vi. Basic principles of non-deductive reasoning, if any

vii. *Synthetic aposteriori* axioms for the special natural and human sciences

Prima facie the problem of justifying a sentential logic is formulating and addressing in the metalanguage the question about the proper system to be used in (i). Formal semantics are "in the right place" for justifying parts of a formal system, viz., they are in the metalanguage. But are they of "the right strength " or do they say "the right thing?" The semantics may require assumptions as strong, if not stronger than the principles accepted in the object language. Of more importance for setting aside the *prima facie* formulation of the problem as the proper formulation is the fact that the metalinguistic statements of the formal semantics do not "say" what is needed to justify the axioms in the object language. The statements of the formal semantics are "is" statements and we need "ought" statements for justification. The metalinguistic claims about provability which I am proposing as the proper claims to make about justification are also "is" claims. However, the provability claims fit in with a general normative claim that people ought to do what they say they will do. When people put a particular system of sentential logic in (i) of the above idealization they are promising to draw inferences in accordance with the rules for that system. The mathematical claims I propose as the genuine logic serve to point out that certain procedures do, or do not, conform to what was promised. This paper, then, is the result of an effort to overcome the *prima facie* formulation of the problem of justifying a logic.

---

[2] My suggestion that the significant metalinguistic claims about logic are those about formal provability has affinities to [Carnap, 1937].

Historically and "logically" the *prima facie* problem arises because of challenges to so-called classical sentential logic PC. PC was the logic presented by Russell and Hilbert. PC is a plausible candidate for sentence logic because it is simple and presents exactly the valid arguments in the clearest sense of validity, viz., those for which it is impossible for the premises to be true and the conclusion false, with only true and false as the truth values. Readers of this volume are familiar with the many challenges made against PC such as: it provides no adequate representation of conditionals, it allows *reductio* existence proofs. Let us focus on the challenge that PC notion of validity is too broad because it warrants as valid inferences in which there is no connection between the premises and conclusion. In particular, PC warrants as valid $A \wedge \neg A \vdash B$ and $B \vdash A \vee \neg A$ where the provability sign $\vdash$ represents logical entailment. Consequently, much of the mid-twentieth century work on developing an improved alternative to PC was developing and justifying systems of sentence logic containing a sentential connective $\rightarrow$ for logical entailment in which $A \wedge \neg A \rightarrow B$ and $B \rightarrow .A \vee \neg A$ are not theorems. Of course many systems have been developed to preserve relevance between premises and conclusion and to prevent derivational chaos in the presence of contradictions. It is this variety which raises so sharply the problem of selecting a correct system.

Many of the objections to PC's "broad" notion of validity could be addressed by simply intervening in applications of logic to say: "We do not want to use that kind of derivation in this context." There is, though, a bias that somehow we should not have to intervene to restrict uses of logic. Intervening to dismiss styles of argument is *ad hoc*. The interventions need not be unprincipled even if *ad hoc*. Nevertheless, these interventions would be subject to dispute and dispute is what we want to avoid with logic. Also for artificial intelligence and reasoning from data banks in which there may be contradictions we should have a system of reasoning which does not go "haywire" and prove everything unless we intervene to stop it. So, I stay committed to the idealization of formalizing principles of inference to be used. In light of that commitment I present system E as an exhibit for how to do "genuine" logic for a system of formal logic.

3 SYSTEM E OF ENTAILMENT

The system E of entailment was developed by A.R. Anderson and Nuel Belnap in 1958 by modifying a system $\prod$ created in 1956 by William Ackermann.

E1. $A \wedge B \rightarrow A$

E2. $A \wedge B \rightarrow B$

E3. $A \rightarrow .A \vee B$

E4. $B \rightarrow .A \vee B$

E5. $A \wedge (B \vee C) \rightarrow .B \vee (A \wedge C)$

E6. $A \to \neg\neg A$

E7. $\neg\neg A \to A$

E8. $((A \to A) \wedge (B \to B) \to C) \to C$

E9. $A \to B \to .B \to C \to .A \to C$ Suffixing

E10. $A \to B \to .C \to A \to .C \to B$ Prefixing

E11. $(A \to A \to B) \to .A \to B$ Contraction

E12. $(A \to B) \wedge (A \to C) \to .A \to .B \wedge C$

E13. $(A \to C) \wedge (B \to C) \to .A \vee B \to C$

E14. $A \to B \to .\neg B \to \neg A$

E15. $A \wedge \neg B \to .\neg(A \to B)$

The rules of proof are MP, $A \to B, A \vdash B$, and Adjunction:

$$A, B \vdash A \wedge B.$$

E-model structures and matrices will be used to show that $A \wedge \neg A \to B, B \to .A \vee \neg A$, and $(A \vee B) \wedge \neg B \to A$ are not theorems. Before presenting a formal semantics for E, let me illustrate my thesis about the practices which constitute "genuine logic" by exhibiting a formal proof within E.

I am proposing that genuine formal sentential logic consists of those practices which we employ to determine with certainty what may and may not be done in particular systems and whether or not procedures of particular systems have been correctly followed. We engage in these practice while using a metalanguage to talk of sentential logic systems. These practices are mathematical because they concern what can or cannot be done with certain structures. They are finitistic because they have the appropriate certainty only when the structures considered are small enough so that we do not have to make many inferences about what else might be in the structure. When using a formal logic a basic claim is a truth claim that such-and-such a sequence is a proof. These are claims of the form $\Gamma \vdash_s \Psi$ when a sequence of wffs. allegedly forming a proof is offered.

Assume Transitivity has been established as a derived rule in E. Consider the following sequence to support $\vdash_E A \to A$

1. $(A \to A) \wedge (B \to B) \to .A \to A$ E1
2. $(A \to A) \wedge (B \to B) \to (A \to A) \to ((A \to A) \to B) \to (A \to A) \wedge (B \to B) \to B)$ Instance of E9
3. $(A \to A) \to B) \to (A \to A) \wedge (B \to B) \to B$ MP1,2
4. $((A \to A) \wedge (B \to B) \to B) \to B$ E8 with B for C

5. $(A \to A) \to B) \to B$ Transitivity 3,4

6. $(A \to A) \to A) \to A$ A for B in 5

7. $(A \to A) \to A =_{df} A$ System definition

8. $A \to A$ 6,7

It is tedious and often tricky to check a proof. But if the proof is not too long, we can tell for certain whether or not it is a proof in the system in question — whether or not the proponents of a system have played by their rules. Strictly speaking the sequence of formulas is not a proof. The sequence is a pattern we display in a proof that the end formula of the sequence is properly linked to the axioms so that it can be called a theorem.

4 ON USE AND INTERPRETATION OF *E*-MODEL STRUCTURES — *Ems*

The purpose of introducing E model structures is to collect reminders that the metalinguistic work with E model structures is insufficient for justifying any system of logic; not even the system which they characterize. Even work with the formal semantics to establish claims about provability and unprovability within the system of sentential logic are far stronger and hence more controversial than the sentential inferences they are supposed to justify. Furthermore efforts to read the postulates for model structures as some type of truths about reasoning are artificial and certainly not beyond controversy. However, introduction of model structures is not irrelevant to my goals for this essay. Model structure work provides heuristics for constructing finitistic justification procedures; especially small finite matrices. The values of these matrices are not plausibly interpreted as truth values. Nonetheless, use of the matrices to show unprovability is the sort of strict finitistic mathematics which I am here proposing as the practices which are genuine formal sentential logic.

Here I should note that showing a matrix is suitable for a system can be finitistic mathematics if mathematical induction is accepted. Mathematical induction is typically finitistic, in my judgment. We show by reference to surveyable structures, taken as one step to the next, that at a next step a counterexample to the theorem will not occur.

An *Ems* $\underline{\text{M}}$ is an ordered 4-tuple: $\langle K, P, R, * \rangle$, where K is a nonempty set of points, P is a unary predicate of Ks, R is a ternary relation among the Ks, and $*$ is a unary operation on the Ks into the Ks, and the following postulates hold for all a, b, c, d etc., in K. A version of the *Ems* postulates is as follows.

E1 $(\exists x)(Px \wedge Rxaa)$

E2 $(\forall x)((Px \wedge Rxab) \supset (\exists y)(Py \wedge Rayb))$

E3 $(\forall x)((Rabx \wedge Rxcd) \supset (\exists y)(Racy \wedge Rbyd))$

E4 $Rabc \supset (\exists x)(Rabx \wedge Rxbc)$

E5 $(\forall x)((Px \wedge Rxab) \wedge Rbcd \supset .Racd)$

E6 $(\forall x)((Px \wedge Rxab) \wedge Rcbd \supset .Rcad)$

E7 $a ** = a$

E8 $Rabc \supset Rac * b*$

E9 $Raa * a$

Presentation of valuation of the postulates on an E model structure provides an occasion for attempting English readings of Px and $Rxyz$. I try to interpret valuations as "data entry" and the postulates as principles of reason for data entry. I do not have a clear idea of what I want to accomplish when I struggle to find an interpretation of model structure postulates as truth claims about some subject matter to which inference drawing is relevant. I am here trying to exhibit the confusion of my thought that if the postulates could be read as truths then using a system of logic they characterized would somehow be acting in accordance with the truth and thereby justified.

A valuation procedure construct sets of formulae Hi for $i\varepsilon K$. Hi stands for the data base formed at i. When there is no attempt to interpret the postulates as truths, Hi could be called simply 'set-ups'. The evaluation rules below continue interpretation of set-ups as data bases. Px tells us: The data base Hx formed at x contains all logical entailments. Call such data bases "origin data bases" or simply "origins." Often origins are designated with o and subscripted o. This reading for Px raises suspicion that any effort to read the postulates as truths justifying the logic characterized by them will "beg the question" What are the logical entailments?

If you are going to draw conclusions from data in a data base which you have under consideration you might look for principles which tell what is entailed by data. The first place to look for such entailment principles is logic itself. $Rabc$ tells us: The data base formed at c contains the consequents of all the entailments in the data base at a whose antecedents are contained in the data base at b. In formal words, $Rabc$ tells us: If $(A \rightarrow B\varepsilon Ha), (A\varepsilon Hb)$, then $(B\varepsilon Hc)$.

A data entry on an Ems is started by making an assignment of each variable of system E at each point in K to get a base of variables Va at each point a in K. An assignment of variables is a function $f()$ on KXV, where V is the set of variables of system E, into $\{1, 0\}$ subject to the following Inclusion Condition.

If $(\exists x)(Px \wedge Rxab)$ and $f(p, a) = 1$, then $f(p, b) = 1$. Given a variable assignment we have variable bases Va at each point Va. $Va = \{p : f(p, a) = 1\}$.

Formulas are now entered into the these bases of variables to get data bases or set-ups at each point a, $a \in K$. The data bases are the smallest sets of formulas developed by the following data entry procedures.

i. If A is a variable, $a \in Ha$ iff. $A \in Va$

ii. $A \wedge B \in Ha$ iff. $A \in Ha$ and $B \in Ha$. $\wedge$**-normality**

iii. $A \vee B \in Ha$ iff. $A \in Ha$ or $B \in Ha$. $\vee$**-normality**

iv. $\neg A \in Ha$ iff. $A \notin Ha*$, where $Ha*$ is the data base formed from variable base $Va*$. **Imperfect Consistency**

v. $A \rightarrow B \in Ha$ iff. $(\forall x)(\forall y)(Raxy \wedge A \in Hx \supset B \in Hy)$, **Triad Rule**

A **formula** A **is valid on an** *Ems.* $\underline{\mathrm{M}}$ if $A \in Hx$ for all x such that Px for all data entries on $\underline{\mathrm{M}}$. A **formula is** E**-valid** if it is valid on all *Ems.*

4.1 Applications of Ems to determine provability in E

Some Lemmas facilitate use of *Ems* in validity arguments. The lemmas are primarily to support a correctness theorem: If A is provable in E, then a is E-valid. The converse stating the *Ems* completeness of E has been proved by R. Routley and R.K Meyer.[3]

LEMMA 1. *If* $(\exists x)Px \wedge Rxab)$ *and* $A \in Ha$*, then* $A \in Hb$*, i.e.,* $Ha \subseteq Hb$.

Proof is by induction on the length of A. Consider a few steps of the proof of what is called "**the General Inclusion Condition**." These steps remind us how postulates, E3 and E8, which play a crucial role in a classical logic argument are suspicious tools for playing a significant role in justifying E as the correct sentential logic.

The inclusion condition for assignment of variables gives the basis. Assume the lemma holds for formulas shorter than A. Focus on two cases: A has the form $\neg B$ or A has form $B \rightarrow C$. Assume for both cases we have a Po_1 and Ro_1ab. If $\neg B \in Ha$, then by Imperfect Consistency, $B \notin Ha*$. From E8, $Ro_1b * a*$ follows from Ro_1ab. Now the induction assumption tells us that $B \notin Hb*$; because if B were in $Hb*$ we would have $B \in Ha*$ from the induction assumption and $Ro_1b{*}a{*}$. Imperfect Consistency now yields $\neg B \in Hb$ from $B \notin Hb*$. For the $B \rightarrow C$ case consider the following *reductio.*

1. $B \rightarrow C \in Ha$ with Ro_1ab
2. $B \rightarrow C \notin Hb$, for *reductio*

From the Triad Rule applied to $B \rightarrow C \notin Hb$, we get an Hc and Hd such that

3a. $B \in Hc$ with $Rbcd$

3b. $C \notin Hd$

From (1) and (3) along with an instantiation of E3 we get:

4. $Ro_1ab \wedge Rbcd$

---

[3] A version of the completeness proof is on found in [Kielkopf, 1977, 317ff].

5. $Ro_1ab \wedge Rbcd \supset (\exists y)(Ro_1cy \wedge Rayd)$. Classical Predicate Logic with (4) and (5) yields (6a) and (6b).

6a. Ro_1cy

6b. $Rayd$.

From the induction assumption on (3a) and (6a) we get:

6. $B \in Hy$.

Now the Triad Rule applied to (1), (6b) and (7) gives:

7. $C \in Hd$.

Since (8) and (3b) are inconsistent we have the $B \rightarrow C$ portion of the induction step by *reductio.*

LEMMA 2. *If* $(B \in HaorC \in Ha)$ *but* $B \notin Ha$, *then* $C \in Ha$.

Call use of Lemma 2 the Elimination argument, EA. We use disjunctive syllogism in our metalanguage even if system E does not have disjunctive syllogism as a valid inference!

LEMMA 3. *If* Px, *then* $A \rightarrow A \in Hx$, *Rule of identity.*

The General Inclusion Condition permits us to develop some fairly short *reductio* arguments about E-validity of an entailment formula $A \rightarrow B$ in the following manner. Assume $A \rightarrow B \notin Ho$ where o is some origin point. By Triad Rule we have that there are points a, b such that $Roab, A \in Ha$ but $B \notin Hb$. However, $Roab$ tells us that $Ha \subseteq Hb$. So we can directly go to consideration of $A \in Hb$ and $B \notin Hb$.

"First degree entailments, fde," are entailment formulas with only one →. The observations of the preceding paragraph show us that when we are testing first degree entailments for E-validity we need only consider two points $\{o, b\}$ and, in effect, use only the 4-value Smiley matrix, introduced below. But if we use only $\{o, b\}$ we need to set $o* = b$ and $b* = o$ which creates a model in which the origin point can contain contradictions. Ems in which no P point o is such that $o* = o$ are called non-normal. For this presentation for fde in Illustrations 6, 7,and 8 I will consider a normal Ems with $\{a, b.c\}$ where Pa and $a* = a, b* = c, c* = b$.

The following illustrations provide reminders for at least two points. The primary point is that the work with formal semantics is really more work with a formal system and not closely connected with justifying any formal system. The secondary point is that formal semantics, which have been proved to characterize a system do provide tableaux techniques for determining provability in the system which are simpler and clearer than construction of formal proofs in the system. I hope the secondary point reminds readers that I am not dismissing the significant intellectual achievement of formal semantics simply because I am dismissing formal semantics as providing a justificatory role. The first two illustrations show steps in proving *Ems* are suitable for *E* while the next three show using *Ems* to show that certain controversial formulas are not *Ems* valid and not *E* theorems under the provable assumption that *Ems* characterize *E*.

ILLUSTRATION 4. $\neg\neg A \to A$ is E-valid. (E axiom E7 is E-valid)

Assume for *reductio* $\neg\neg A \to A \notin Ho$ for some Ho and $Roab$

1. (a) $\neg\neg A \in Hb$
   (b) $A \notin Hb$
2. $\neg A \notin Hb*$ Imperfect Consistency on (1a)
3. $A \in Hb**$ Imperfect Consistency on (1b)
4. $A \in Hb$ Use of *Ems.* postulate E7 on 3.

The *reductio* proof is complete because 4 and (1b) would be a contradiction for us in our talk about data bases or set-ups.

ILLUSTRATION 5. $A \to B \to .B \to C \to .A \to C$, i.e., E9 (suffixing) is E-valid.

Assume for *reductio*

1. (a) $A \to B \in Hb$
   (b) $B \to C \to .A \to C \notin Hb$, with $Po, Roab$

The Triad Rule applied to (1b) gives us:

2. (a) $B \to C \in Hc$
   (b) $A \to C \notin Hd$ with $Rbcd$

The Triad Rule applied to (2b) gives us:

3. (a) $A \in He$
   (b) $C \notin Hf$ with $Rdef$

Now digress to a "little exercise" in classical predicate logic.

i. $(\forall x)(\forall z)(\forall u)(\forall v)(\forall w)(Rbcx \wedge Rxvw) \supset (\exists y)(Rbvy \wedge Rcyw))))))$

E3 with explicit universal quantification.

ii. $Rbcd \wedge Rdef$ From (2b) and (3b)

iii. $Rbcd \wedge Rdef) \supset (\exists y)(Rbey \wedge Rcyf)$ Universal Instantiation on i.

iv. $(\exists y)(Rbey \wedge Rcyf)$ MP for $\supset$ on ii and iii

v. $Rbey \wedge Rcyf$ Existential Instantiation on iv

The Triad Rule with Rbey, (1a), (3a) gives

4. $B \in Hy$

The Triad Rule with $Rcyf$, (2a), (4) gives us:

5. $C \in Hf$

Since (5) and (3b) contradict one another, the *reductio* is complete.

The E-validation of E10 requires even more predicate logic manipulation of the E postulates than did validation of E9. The need to turn to undecidable predicate logic to validate formulas should make us suspicious about use of Ems as a decision procedure. Alasdair Urquhart proved the undecidability of system E in 1984.[4] Here is a place for a reminder to make one suspicious about the Ems. postulates as logic justifying truths. The reminder is their use in proofs to show that system E is characterized by Ems validity. When working through the proofs it is easy to get the impression that the postulates are selected to get model structures to characterize system E or whatever system is being considered. For instance, alternative relevance logics such as system R, which has $A \rightarrow .A \rightarrow .A \rightarrow A$ as theorem, is shown to be characterized by model structures to which we add the postulate: $Rabc \supset Rbac$.

As we shall see the following tableaux style invalidity arguments correspond nicely to getting counterexamples on a matrix.

ILLUSTRATION 6. $A \wedge \neg A \rightarrow B$ is not E-valid

Consider an $Ems : K = \{a, b, c\}, Pa, a* = a, b* = c, c* = b$ and R specified by the postulates. We re-label a as o. Postulate E1 tells us $Robb$. Assume $A \wedge \neg A \rightarrow B \notin Ho$

So we have:

1. (a) $A \wedge \neg A \in Hb$
   (b) $B \notin Hb$
2. (a) $A \in Hb$
   (b) $\neg A \in Hb$$\wedge$-normality on (1a).

Step (2b) puts a contradiction in data base Hb. But a contradiction in a data base we are talking about is not a contradiction for us.

3. $A \notin Hb*$ Imperfect Consistency on (2b).

With $b* = c$, (3) gives us:

5. $A \notin Hc$

The removal of all logical signs from all parts of the formula $A \wedge \neg A \rightarrow B$ without contradicting ourselves shows that we can assume that it is not E-valid.

Here it is interesting to note that corresponding to the inconsistency in Hb we have an incompleteness in Hc since neither A nor $\neg A$ are in Hc. $\neg A \in Hc$ iff. $A \notin Hc*$ which is Hb. Since, (2a) tells us $A \in Hb$ we have:

---

[4] [Urquhart, 1984, 1059–1073.]

5. $\neg A \notin Hc$.

As we shall see in the matrix section, this analysis of $A \wedge \neg A \rightarrow B$ to its basic parts shows us that we have assigned A a value 110 or 010 on an 8-value world line matrix while B has been assigned a value 101, 100, 001 or 000 . A look at the 8-value $\wedge$ charts below shows $A \wedge \neg A$ get -1 for these values and the Smiley 8-value $\rightarrow$ table shows $-1 \rightarrow$ any of the four values for B get -3.

ILLUSTRATION 7. $B \rightarrow .A \vee \quad \neg A$ is not E-valid,

Assume $B \rightarrow .A \vee \neg A \notin Ho$ in an *Ems.* with same structure as that for Illustration 6.

1. (a) $B \in Hb$
   (b) $A \vee \neg A \notin Hb$

$\vee$-normality tells us:

2. (a) $A \notin Hb$
   (b) $\neg A \notin Hb$

Imperfect Consistency on (2b) tells us:

3. $A \in Hc$

Since we have reduced the formula to simplest parts without contradicting ourselves we have shown that $B \rightarrow .A \vee \neg A$ need not hold at a P-point in an *Ems.*

Again it is interesting to note that corresponding to the incompleteness of Hb exposed in (2) we have an inconsistency in Hb's $*$ image Hc because (2a) tells us $\neg A \in Hc$.

ILLUSTRATION 8. $(A \vee B) \wedge \neg B \rightarrow A$ is not *Ems.* valid.

Assume same *Ems* structure as in previous two illustrations.

1. Assume $(A \vee B) \wedge \neg B \rightarrow A \notin Ho$ (where $o = a$ and Pa)

2. (a) $A \vee B \in Hb$ $\wedge$-normality and triad rules on 1
   (b) $\neg B \in Hb$
   (c) $A \notin Hb$

To apply $\vee$-normality to 2a we cannot set $A \in Hb$ because that would be a contradiction for us with 2c. However, we can put a contradiction into Hb by using $\vee$-normality on 2a to set

3. $B \in Hb$.

4.2 Dismissal of efforts to read the E postulates as fundamental truths

Can the postulates be read as some kind of truths about data bases? Consider a few efforts to read the postulates in light of the above suggestions that the postulates are about data bases. What might E1 tell us?

E1: $(\exists x)(Px \wedge Rxaa)$

E1 says more than $Ha \subseteq Ha$. There is a data base at x containing all logical entailments such that for any data base entered at point a if $A \rightarrow B$ is a logical entailment and A belongs to the data base entered at point a B will also belong to the data base entered at point a. I.e., all data bases are closed under logical entailment. Is this true? Does reason require us to close data basis under entailment if we are to be rational?

E2: $(\forall x)((Px \wedge Rxab) \supset (\exists y)(Py \wedge Rayb))$

If $Ha \subseteq Hb$ then if an item in Ha entails a logical entailment LE, $LE \in Hb$. I do want not to continue trying to read E2, or any of the *Ems* postulates, in persuasive ways. These are efforts to try to lead people to accept complicated claims as plausible rather than pointing out fundamental truths. These efforts are unpleasant because I have no sense of discovering truth or even “inventing” an interesting theory; only a sense of pleading. (My original plan for this essay was to present such a reading as a case for the *Ems* Postulates.) Here I want reading the postulates to go far enough to remind us that the postulates are no more persuasive about what is logically correct than are the axioms read as claims about entailment. I do not think any reading of these postulates will show that system E which is characterized by validity on *Ems* is the correct logic. Reading the postulates about model structures is reading them as norms or truths about a subject matter which is somehow the subject matter of logic. If the reading showed E to be the correct logic, the postulates as read should be self-evident norms or necessary truths about this subject matter of logic. I dismiss the suggestion of nontrivial self-evident norms. If they were contingent truths their truth would be contingent upon the type of data bases discussed and would not be general enough for logic. If contingent truths they should be below logic on the “schema for science.” If they were necessary truths, they would have to be analytic or synthetic apriori. An attempt to read them will, I think, show them to be non-analytic. Some are too hard to understand from the meaning of the terms. Others present, in effect, controversial claims such as double negation. If they were synthetic apriori, they are not the right kind to justify a logic. One needs to give too much argument to make a case for accepting them when alternatives are proposed. Logic, which is supposed to be used to settle arguments, should not require argument — at least not too much argument.

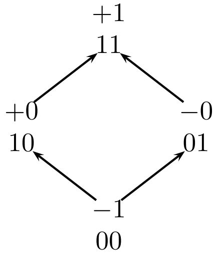
However, I wish to conclude this paper by showing that *Ems* can be guides for developing finite matrices. Once we have matrices we can use them to decide finitistically what is, and what is not provable in a system. Of course, the purpose of the discussion of matrices is not to inform readers about algebraic semantics. The purpose is to remind us of the finitistic character of work with particular matrices to which the non-finitistic metamathematical work with formal semantics can be a guide.

5 ON WORLD-LINE MATRICES FOR TWO AND THREE "POINTS"

The values on the matrices presented are obtained the ways of assigning $\{1, 0\}$ to a series of points or worlds. If N is the number of points, the number of matrix. values is 2^N. If there are 2 points $\{a, b\}$ in the order $\langle a, b \rangle$ the matrix values are $+1, +0, -0, -1$ as follows.

	A	B
+1	1	1
+0	1	0
-1	0	1
-0	0	0

These values can be charted on the 4-value diamond lattice



A **world line matrix for an entailment** system with sentential connectives $\wedge, \vee, \neg, \rightarrow$ is an ordered sextuple $\langle B, D, \wedge, \vee, \rightarrow, \neg \rangle$ where B is a set world line values, D is a proper subset of the B considered as the designated values $\wedge, \vee$. $\rightarrow$ are binary operations on B into B while $\neg$ is a unary operation on B into B for representing the sentential connectives.

Tables show how the matrix operations represent the sentential connectives for two worlds or points.

The rules for evaluating $\wedge$ and $\vee$ are on the tables below.

A/B means A down, B across. To compute $A \wedge B$ use diamond diagram to find where values for A and B meet below and to compute value for $A \vee B$ find where values of A and B join above.

If Ψ is a formula, use $V(\Psi)$ to show the world matrix value of Ψ. The $\wedge$ table above shows $V(A \wedge B) = +0$ when $V(A) = +1$ and $V(B) = +0$.

$\wedge$ A/B	+1 11	+0 10	-0 01	-1 00	/	$\vee$ A/B	+1 11	+0 10	-0 01	-1 00
+1 11	+1	+0	-0	-1		+1 11	+1	+1	+1	+1
+0 10	+0	+0	-1	-1		+0 10	+1	+1	+1	+0
-0 01	-0	-1	-0	-1		-0 01	+1	+1	-0	-0
-1 00	-1	-1	-1	-1		-1 00	+1	+0	-0	-1

The rules for $\wedge$ and $\vee$ are basically the same for all world line matrices. The values form a lattice and the value for $\wedge$ is where values "join below" while values for $\vee$ are where values "meet above." Rules for $\rightarrow$ and $\neg$ vary. One set of rules for evaluating $\rightarrow$ and $\neg$ are on the tables below.

$\rightarrow$ A/B	+1, 11	+0, 10	-0, 01	-1, 00		A	$\neg A$
+1, 11	+1	-1	-1	-1		+1, 11	-1
+0, 10	+1	+1	-1	-1		+0, 10	+0
-0, 01	+1	-1	+1	-1		-0, 01	-0
-1, 00	+1	+1	+1	+1		-1, 00	+1

To compute the value for $A \rightarrow B$ give it +1 if value of A is included in Value of B; otherwise give it -1. If arrows on 4-value diamond lattice lead from value of A to value of B, value of A is included in value of B.

Rationalizing the rule for $\neg$ requires using the notion of a point or world being the negation complement of another. For world line matrices there is an operation $*$ which maps the set of points into itself. Each point has it's $a*$. For normal negation $a* = a$ for all points a. For negation it is useful to have a more complex notation specifying the value of a formula A at each point $a : v(A, a) = n$ where n is 1 or 0. For instance, when the value of A is +0, viz., $V(A) = +0$, then $v(A, a) = 1$ and $v(A, b) = 0$ if we have two worlds or points $\{a, b\}$. If x is a world or point, $v(\neg A, x) = 1$ if $v(A, x*) = 0$ and $v(\neg A, x) = 0$ if $v(A, x*) = 1$. Note: if $x = x*, v(\neg A, x) = 1$ iff. $V(A, x) = 0$.

The basic procedure is to evaluate $\neg A$ in the column under point a by looking at the $a*$ column and putting in the opposite value. See the next chart where $a* = b$ and $b* = a$. To get the values in the a column for $\neg A$ take opposites of the values in the b column for A; to get values in the b column for $\neg A$ take opposites of the values in the a column for A. Note columns for $A \wedge \neg A$ and $A \vee \neg A$.

A	A	b		$\neg A$	a	b		$A \wedge \neg A$	$A \vee \neg A$	
+1	1	1		-1	0	0		-1	+1	
+0	1	0		+0	1	0		+0	+0	
-0	0	1		-0	0	1		-0	-0	
-1	0	0		+1	1	1		-1	+1	

If we specify $+1, +0$ as the designated values the tables for $\wedge, \vee, \neg$ and $\rightarrow$ give the Timothy Smiley matrix which was shown in [1975] to be characteristic for the first degree fragment of system E.[5]

A *formula is valid on a matrix* if its value is a designated value for all values of its components. A *matrix is suitable for a system* if all theorems of the systems are valid on the matrix. A *matrix is characteristic* for a system if it is suitable for the system and all formulas valid on the matrix are theorems of the system.

A 1940 result of J. Dugundji established that no reasonably strong nonclassical logic such as most modal logics and system E have no finite characteristic matrix.

Here we can use the Smiley matrix to show the following.

$A \wedge \neg A \rightarrow B$ is not a theorem of E. (Nice result)

Set $V(A) = +0$ and $V(B) = -1$. Now $V(A \wedge \neg A) = +0$. Hence, $V(A \wedge \neg A \rightarrow B) = +0 \rightarrow -1 = -1$

$B \rightarrow .A \vee \neg A$ is not a theorem of E. (Nice result)

Set $V(A) = +0$ but $V(B) = +1$. Now $V(A \vee \neg A) = +0$. Hence, $V(B \rightarrow .A \vee \neg A) = +1 \rightarrow +0 = -1$

$(A \vee B) \wedge \neg B \rightarrow A$ is not a theorem. (Embarrassing result)

Set $V(A) = +0$ and $V(B) = -0$. Now $V(A \vee B) = +1, V(\neg B) = -0$ giving $V(A \vee B) \wedge \neg B) = +1 \wedge -0 = -0$. So V for whole formula is $-0 \rightarrow +0 = -1$.

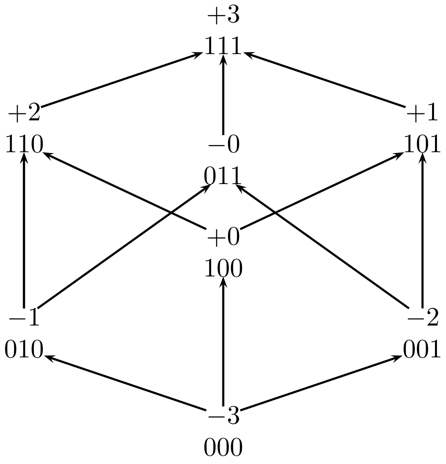
A goal is to appreciate model structure semantics as a heuristic for constructing matrices. E model structures are based on a triadic relation $Rabc$ between points or worlds. This triadic relation “naturally” suggests using at least three points or worlds. So it is interesting to consider three point world line matrices. If there are 3 points {a,b,c} we have the eight values, +3,+2,+1,+0, -0,-1,-2,-3 as follows.

The 8 values of a 3 World line matrix form a cube lattice.

---

[5] [Anderson and Belnap, 1975].

	a	B	c
+3	1	1	1
+2	1	1	0
+1	1	0	1
+0	1	0	0
-0	0	1	1
-1	0	1	0
-2	0	0	1
-3	0	0	0



The cube lattice provides the following matrix tables for ∧ and ∨. I present them because matrices "ought to have arrays of numbers."

For eight value ∧ table use cube lattice diagram and trace arrows back from each value until you reach the 1^{st} point at which they meet.

For eight value ∨ table use cube lattice diagram and trace arrows up from each value until you reach the 1^{st} point at which they join.

Consider a ¬ table where we have $a* = a, b* = c$ and $c* = b$. Informally the a column for $\neg A$ negates the a column for A while the b column for $\neg A$ negates the c column for A and the c column for $\neg A$ negates the b column for A.

Consider a table for $A \rightarrow B$ using rule $V(A \rightarrow B) = +3$ if $V(A)$ included in $V(B)$ otherwise it is -3. If an arrow path goes from $V(A)$ to $V(B), V(A)$ is included in $V(B)$. Call it Smiley 8-value → table.

The Smiley 8-value → table is too generous to invalidate significant formulas which are not E-theorems such as $B \rightarrow .A \rightarrow A$. $B \rightarrow .A \rightarrow A$ should not be a theorem in any relevance preserving logic.

∧	+3 111	+2 110	+1 101	+0 100	-0 011	-1 010	-2 001	-3 000
+3 111	+3	+2	+1	+0	-0	-1	-2	-3
+2 110	±2	+2	-3	-3	-1	-1	-3	-3
+1 101	+1	-3	+1	+0	-2	-3	-2	-3
+0 100	+0	+0	+0	+0	-3	-3	-3	-3
-0 011	-0	-1	-2	-3	-0	-1	-2	-3
-1 010	-1	-1	-3	-3	-1	-1	-3	-3
-2 001	-2	-3	-2	-3	-2	-3	-2	-3
-3 000	-3	-3	-3	-3	-3	-3	-3	-3

∨	+3 111	+2 110	+1 101	+0 100	-0 011	-1 010	-2 001	-3 000
+3 111	+3	+3	+3	+3	+3	+3	+3	+3
+2 110	+3	+2	+3	+2	+3	+2	+3	+2
+1 101	+3	+3	+1	+1	+3	+3	+1	+1
+0 100	+3	+2	+1	+0	+3	+2	+3	+0
-0 011	+3	+3	+3	+3	-0	-0	-0	-0
-1 010	+3	+2	+3	+2	-0	-1	-0	-1
-2 001	+3	+3	+1	+1	-0	-0	-2	-2
-3 000	+3	+2	+1	+0	-0	-1	-2	-3

A	a	b	c	$\neg A$	A	b	c	$A \wedge \neg A$
+3 111	1	1	1	-3	0	0	0	-3
+2 110	1	1	0	-1	0	1	0	-1
+1 101	1	0	1	-2	0	0	1	-2
+0 100	1	0	0	-0	0	1	1	-3
-0 011	0	1	1	+0	1	0	0	-3
-1 010	0	1	0	+2	1	1	0	-1
-2 001	0	0	1	+1	1	0	1	-2
-3 000	0	0	0	+3	1	1	1	-3

$\rightarrow$	+3 111	+2 110	+1 101	+0 100	-0 011	-1 010	-2 001	-3 000
+3 111	+3	-3	-3	-3	-3	-3	-3	-3
+2 110	+3	+3	-3	-3	-3	-3	-3	-3
+1 101	+3	-3	+3	-3	-3	-3	-3	-3
+0 100	+3	+3	+3	+3	-3	-3	-3	-3
-0 011	+3	-3	-3	-3	+3	-3	-3	-3
-1 010	+3	+3	-3	-3	+3	+3	-3	-3
-2 001	+3	-3	+3	-3	+3	-3	+3	-3
-3 000	+3	+3	+3	+3	+3	+3	+3	+3

To get an → table closer to validating only E theorems I use the triadic $Rabc$ relation of E model structures the next section. With three points $\{a, b, c\}$ the K8-value → table is based on the following specifications for R: These R specifications were obtained by figuring out what the E model structure postulates required for $\{a, b, c\}$ with $Pa, a* = a, b* = c, c* = b$.[6] Perhaps reading the axioms to determine what is required to specify R is more an exercise in reading logical formulas than mathematics. But once a specification for P and R is obtained constructing a table for → is a paradigm of the type of strict finitistic mathematics I am proposing as the core of logic. I will close, then, with presenting and commenting a bit on how I constructed the → for the following specification of R. And show $B \to .A \to A$ is not valid.

$$Raaa, Rabb, Racc$$
$$Rbbb, Rbab, Rbcb, Rbcc$$
$$Rccc, Rcac, Rcbc, Rcbb$$

→ A/B	+3 111	+2 110	+1 101	+0 100	-0 011	-1 010	-2 001	-3 000
+3 111	+3	-3	-3	-3	-0	-3	-3	-3
+2 110	+3	+2	-3	-3	-0	-1	-0	-3
+1 101	+3	-3	+1	-3	-0	-1	-2	-3
+0 100	+3	+2	+1	+0	-0	-1	-2	-3
-0 011	+3	-3	-3	-3	+3	-3	-3	-3
-1 010	+3	+2	-3	-3	+3	+2	-2	-3
-2 001	+3	-3	+1	-3	+3	-3	+1	-3
-3 000	+3	+3	+3	+3	+3	+3	+3	+3

The program for calculating the values is complicated. The basic principle is $v(A \to B, a) = 0$ iff. $Rabc, v(A, b) = 1$ and $v(B, c) = 0$, which comes from a "negative" version of the Triad Rule. If someone developed a computer program for these computations, as did N. Belnap, the programming is the finistic mathematics.

Heuristic account of calculation for K 8-value → table

---

[6] The R specifications for the Smiley 8-value → are $Raaa, Rabb, Racc, Rbbb, Rbaa, Rbcc, Rccc, Rcba, Rcbb$.

The goal is to determine how to put 1 or 0 in each of the three a, b, c, places: 1^{st} place, 2^{nd} place, 3^{rd} place. Here "drop" means that the value for A in a place goes from 1 to a value of 0 for B in the place specified. There will be three groups of instructions: Those for $Ra\_\_$ to determine whether $A \rightarrow B \in Ha$, those for $R\_b\_$ to determine whether $A \rightarrow B \in Hb$ and those for $R\_\_c$ to determine whether $A \rightarrow B \in Hc$.

Put 0 in 1^{st} place iff.

Drop in 1^{st} place *Raaa* (eg. 111 → 011)
or
Drop in 2^{nd} place *Rabb* (eg. 111 → 101)
or
Drop in 3^{rd} place *Racc*. (eg. 111 → 110)

Put 0 in 2^{nd} place iff.

Drop in 2^{nd} place *Rbbb* (eg. 011 → 101)
or
Drop in 3^{rd} place *Rbcc* (eg. 011 → 010)
or
Drop from 1^{st} to 2^{nd} *Rbab* (eg. 111 → 101)
or
Drop from 3^{rd} to 2^{nd} *Rbcb*. (eg. 001 → 101).

Let me comment on this example. The Triad Rule with *Rbcb* tells us that $A \rightarrow B \notin Hb$ if $A \in Hc$ but $B \notin Hb$. The antecedent 001 for A says $A \in Hc$ while the consequent value 101 for B shows $B \notin Hb$.

Put 0 in 3^{rd} place iff.

Drop in 2^{nd} place *Rcbb* (eg. 010 → 001)
or
Drop in 3^{rd} place *Rccc* (eg. 011 → 110)
or
Drop from 1^{st} to 3^{rd} *Rcac* (eg. 110 → 110)
or
Drop from 2^{nd} to 3^{rd} *Rcbc*. (eg. 010 → 110)

An example: To compute $+3 \rightarrow -0$, i.e., $111 \rightarrow 011$ put 0 in 1^{st} place because a drop in 1^{st} place but 1 in the next two places because none of for putting in 0 are met. Result is 011 or -0.

Assuming that it has been shown that K8 is suitable for E with positive values as designated, it shows that $B \rightarrow .A \rightarrow A$ is not a theorem. Set $V(B) = +3$ and $V(A) = +2$ and get result of -3.

BIBLIOGRAPHY

[Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap, Jr. *The Logic of Relevance and Entailment*, Vol. 1. Princeton: Princeton University Press. 1975.

[Carnap, 1937] R. Carnap. *The Logical Syntax of Language*. Trans. Amethe Smeaton, Countess von Zeppelin. London: Kegan Paul Trench, Trubner & Co. 1937.

[Kielkopf, 1977] C. F. Kielkopf. *Formal Sentential Entailment*. Washington, DC: University Press of America. 1977.

[Urquhart, 1984] A. Urquhart. The undecidability of entailment and relevant implication. *Journal of Symbolic Logic*, 49, 1984, 1059-1073. 1984.

A BRIEF HISTORY OF TRUTH

Stewart Candlish and Nic Damnjanovic

BRIEF INTRODUCTION

Broadly speaking, there are three traditions in recent philosophical writing about truth. First, there is a highly technical literature of interest principally to formal logicians. Secondly, there is the literature of the so-called "Continental" tradition, of consuming interest to those for whom obscurity is a reliable mark of profundity. Thirdly, there is a recognizably interrelated set of writings which give rise to, and develop, the mainstream work on the topic in English (though these writings are not themselves always in English). It is more than enough work for an article such as this to concentrate, as we have decided to do, upon the third — and most accessible — of these traditions, which in any case overlaps the others.

Pascal Engel begins his recent book *Truth* by remarking, 'Truth is a central philosophical notion, perhaps *the* central one. Many other important philosophical notions depend upon it or are closely tied to it ... ' [Engel, 2002, 1]. He goes on to give so many examples that someone who proposes to write about truth might despair at the outset, faced not only with the seeming obligation to say something about everything but also with the suspicion that truth poses a problem of first philosophy: that, as Michael Dummett once suggested about logic, if we do not get it right, we shall get nothing else right. But its very ubiquity might make one suspicious of its importance, on the grounds that it may just be appearing as a proxy for a range of other, or local, topics. And indeed, one of the main issues in recent discussion has been just how much of what has commonly been assumed to be central to the investigation of truth turns out to be really a matter of indifference.

So, for example, among the questions we shall be canvassing are the following. Are the disputes over the nature of truth, and the realism/anti-realism debate, mutually relevant or not? Is the theory of meaning relevant to the nature of truth? Do ontological claims about truthmakers support some accounts of truth over others? Does the plausible suggestion that truth is a norm of assertion require us to order rival theories in terms of their respective capacities for giving an account of it? Does a theory of truth require a theory of justification?

One way in which to handle such questions is directly, taxonomizing the various positions and attempting to map the relations between them. We have chosen, instead, to approach the subject matter historically, fulfilling our editorial brief

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

by tracing the course of the debates as they developed from the turn of the twentieth century to the turn of the twenty-first. This choice is the result of several commitments that we share. One is a preference for looking at the theories of real philosophers rather than the abstractions (and sometimes straw men) of the textbooks, even though this has meant that we have had to make what may seem invidious and idiosyncratic decisions about who is to be a focus of attention, and who ignored or marginalized. Another is the belief that a theory, together with its strengths and weaknesses, is best understood through a developmental account, for the views of philosophers emerge in response to the various competing demands made in actual debate. Yet another is the conviction that bad (or, perhaps worse, time-wasting) philosophy results from amnesia. The history of analytic philosophy can, as we shall see, look like a tale of increasingly elaborate re-invention of a range of rectangular wheels. Part of the story we have to tell is one of how often the taking of sides about truth involves a return to a position already mapped decades previously.

1 EARLY VIEWS

1.1 Introductory comments

The various claims about truth which were being advocated around the start of the twentieth century were largely by-products of their proponents' commitments in metaphysics, the philosophy of logic, and the philosophy of language (though at the time this last branch of philosophy was not the self-conscious sub-discipline it subsequently became). In particular, theories of truth were inextricably entangled with what were then still often called theories of judgement. Among these views we find some which still look familiar, but others which only philosophers with an interest in the history of the period would recognize — although, as so often in philosophy, it is just these half-forgotten and more recondite theories which have, at the time of writing, most recently reappeared in modern dress.

1.2 Primitivism: Moore, Russell

The close of the nineteenth century witnessed an intellectual rupture which was to have momentous consequences for the way in which philosophy subsequently developed. It began with G. E. Moore, who in working on Kant had come to the view that a basic presupposition of any form of idealism (variants of which had dominated Britain's major universities for some decades), that the objects of human knowledge are in some way mind-dependent, had to be rejected. Moore reacted with the zeal of the convert, abandoning idealism for an extreme realism, rejecting the ontological monism which frequently accompanied absolute idealism in favour of an extreme atomism, and, more significantly, persuading Bertrand Russell to follow him.

Moore's resultant ontology is difficult even to comprehend, let alone accept. It begins to emerge in his theories of propositions and truth:

> A proposition is a synthesis of concepts; and just as concepts are themselves immutably what they are, so they stand in infinite relations to one another equally immutable. A proposition is constituted by any number of concepts, together with a specific relation between them, and according to the nature of this relation the proposition may be either true or false. What kind of relation makes a proposition true, what false, cannot be further defined, but must be immediately recognised.
>
> [Moore, 1899, 5]

All this may seem consistent with a correspondence theory of truth, but it quickly becomes clear that Moore does not think of the relation as 'making' the proposition true (or false) by arranging the proposition's concepts to reflect (or not) a corresponding arrangement of existents. One of his reasons, soon to become familiar in the framing of theories of truth, is the difficulty on the correspondence account of making sense of arithmetical truths. But his main reason is far more idiosyncratic. Clearly inspired by the residual influence of idealism, Moore contends that meanings cannot be abstracted from the actual things in the world that are meant. (Moore's conception of meaning involves rejection of distinctions such as that between sense and referent.) Consistent with this opposition to abstraction, and because he holds that meanings are concepts, he goes on to proclaim that 'the world is formed of concepts'. In consequence, he rejects the correspondence theory, maintaining that a judgement's 'truth or falsehood cannot depend on its relation to anything else whatever, reality, for instance, or the world in space and time',[1] and that, while the correspondence theory tries to define truth in terms of existence, any definition should go in the other direction. Since, on this view, the world cannot be distinguished from the totality of propositions, Moore risks losing the distinction between truth and falsehood. At first it seems that he protects himself from this consequence by stipulating that truth is one kind of relation amongst concepts and falsehood another. This view would face the serious objection that it requires both truth and falsehood to be, each of them, a potentially infinite number of kinds of relations because of the potentially infinite complexity of propositions. But Moore quickly moves to a different protective strategy: he adds truth and falsehood to the stock of basic properties. (The move occurs seemingly unawares; this may be the first, inchoate, appearance of Moore's later idea of supervenient properties.) The idea is a version of what is now generally called 'primitivism': truth is a fundamental, indefinable, irreducible property of propositions. One of Moore's arguments for his claim that truth is primitive develops a theme which was also destined for familiarity: that any proposition which purports to define truth must, if the definition is to be correct, itself be true, so that any such definition is bound to be viciously circular.

---

[1] Moore 1899: 8, 18 respectively.

Moore's view, which he soon abandoned, appears little more than an historical curiosity. Its importance, in hindsight, resides partly in its prefiguring of later concerns, but mainly in its impact on Russell, whose own version of this theory is highly significant because of its generation of an extended series of influential reactions to its internal difficulties. Russell's version begins with a sketchy account of judgement which may be called the 'binary theory': judgement is a single primitive binary relation between two entities, a judging mind and a proposition. But whereas Moore had taken the eternal constituents of propositions and put them into the world, Russell started with objects at least many of which are commonplace and constructed propositions from them. A proposition does not consist of words; 'it contains the entities indicated by words'.[2] These Russell called 'terms', and they include, e.g., men, chimaeras, and relations.[3] [From now on, we shall call this *the doctrine of real propositional constituents*.] The reason Russell believes that constituents of propositions are the things the propositions are *about* appears to be a view which emerges explicitly only later in his writings: that the sole alternative is to regard them as ideas, which are 'constituents of the mind of the person judging' and 'a veil between us and outside things'.[4] 'Every term', he says, '... is a logical subject ... possessed of all the properties commonly assigned to substances'.[5] This idea that everything is at bottom an object, and of the same sort, is, Russell thinks, unavoidable: the attempt to deny it leads to self-contradiction.[6] His explanation of the contradiction is unclear, but it looks to be a version of Frege's notorious problem concerning the concept *horse*,[7] namely that if one regards the proposition as composed of both saturated and unsaturated elements (in Frege's vocabulary, of objects and concepts), then it is impossible to talk about the unsaturated ones, for as soon as one puts the unsaturated, predicative, element into subject position it becomes something else, something saturated. This not only makes it impossible to talk about concepts, but certainly looks inconsistent. Had Russell used Fregean terminology, he would have held the constituents of propositions to be, all of them, saturated.

Why did Russell think that propositions, as well as being composed of entities, are themselves entities? Because he held that they were unities,[8] and he subscribed to the principle *ens et unum convertuntur*;[9] in addition, it soon became clear that his early attempts to prove the so-called 'axiom of infinity' require the assumption that propositions are entities[10] — without such a proof, he would have been forced to admit that the theorems of mathematics cannot be derived solely from principles which are true by logic alone. (That this assumption generates paradox is another story.) What makes a proposition a unity? His answer is that its constituents are related by the proposition's verb: 'the true logical verb in a proposition may be always regarded as asserting a relation'.[11] Moreover, the verb, Russell says, 'when used as a verb, embodies the unity of the proposition'.[12]

---

[2] [Russell, 1903, §51].

[3] ibid. §47.

[4] [Russell, 1911, 155].

[5] [Russell, 1903, §47].

[6] ibid. §49.

[7] [Frege, 1892, 45].

[8] [Russell, 1903, §54].

[9] ibid. §47.

[10] [Russell, 1904b].

What, then, is the unity of the proposition? It is what distinguishes a proposition from a list of its constituents, so that unlike a mere list it 'holds together' and says something. But this seemingly undeniable unity, when combined with Russell's principle that 'Every constituent of every proposition must, on pain of self-contradiction, be capable of being made a logical subject',[13] generates a problem. On pain of contradiction, the verb must itself be a term, something capable of appearing as a logical subject, i.e. saturated. But it must be a very unusual kind of term, for while itself being one of the related items it must simultaneously be unsaturated too, the source of the proposition's unity, relating all its constituents. That is, the verb is unlike other terms in that it has, he says, a 'twofold nature... as actual verb and verbal noun, [which] may be expressed... as the difference between a relation in itself and a relation actually relating'.[14] Yet as soon as we make the verb a logical subject, we are forced to identify it as 'a relation in itself' rather than as 'a relation actually relating', destroying the unity of the original proposition in which it was the source of that unity. He illustrates the point like this:

> Consider, for example, the proposition "A differs from B." The constituents of this proposition, if we analyze it, appear to be only A, difference, B. Yet these constituents, thus placed side by side, do not reconstitute the proposition. The difference which occurs in the proposition actually relates A and B, whereas the difference after analysis is a notion which has no connection with A and B.... A proposition, in fact, is essentially a unity, and when analysis has destroyed the unity, no enumeration of constituents will restore the proposition. The verb, when used as a verb, embodies the unity of the proposition,

---

[11] [Russell, 1903, §53]. He gets around the apparent exceptions posed by intransitive verbs like 'breathes' by claiming that in such cases the verb expresses a complex notion which 'usually asserts a definite relation to an indefinite relatum' (ibid. §48).

[12] ibid. §54. Russell talks indifferently of 'verbs' whether he means words or the 'entities indicated by words', i.e. terms. (This explains the frequency of subsequent accusations of use/mention confusions.) The indifference results from his inclination to the view that English grammar gives us — by and large — a transparent window through which to view reality. Although his inclination was temporary, this fantasy about grammar is surprisingly persistent, though generally as one or other unexamined presupposition rather than, as here, a doctrine explicitly embraced. It recurs with numbing frequency in the debate over truth, generally in the form that, because ' ... is true' is a grammatical predicate we should, *prima facie*, expect truth to be a property. (A recent example is [Horwich, 1998a, 37].) Some effort is made at exposing the fantasy in [Oliver, 1999] and [Candlish, 2001].

[13] [Russell, 1903, §52].

[14] ibid. §54.

> and is thus distinguishable from the verb considered as a term, though I do not know how to give a clear account of the precise nature of the distinction.
>
> [Russell, 1903, §54]

Russell's problem, then, is that while he cannot deny propositional unity, he can find no account of the proposition which can do justice to it. Perhaps anxious to get on with mathematical matters, he left the matter unresolved. Opinions differ over how serious that problem is.[15] But a related difficulty is certainly serious: whether true or false, a proposition is a unity, hence on Russell's view an entity. In fact it is a complex entity whose constituents are the things it is about, which makes it hard to see how it can differ from what in his later vocabulary would be called a fact. The difficulty, in thin disguise, is just the perennial conundrum: how is false judgement possible? The source of the difficulty is the combination of Russell's attachment both to the unity of the proposition and to the doctrine of real propositional constituents. This makes it hard for him to give a sensible account of truth, and the correspondence theory is noticeably absent from *The Principles of Mathematics.* Rather, he turns to primitivism, saying merely that truth is an unanalysable property: true propositions just have it, false ones just lack it.[16] The world, then, contains both objective falsehoods and objective truths: 'objective', here, meaning that they are entities in no sense mind-dependent.

Some interpret these early forms of primitivism as a version of what has subsequently come to be called the identity theory of truth, according to which truth consists in an identity between truth-bearer and truthmaker.[17] This vacillation in recent commentary between the identity theory and primitivism is prefigured in Meinong, at least as described by Findlay.[18] Meinong's 'objectives', such as the being-white-of-snow and the being-an-integer-between-3-and-4, divide into two sorts, the factual (*tatsächlich*) and the unfactual (*untatsächlich*). There are no entities between our minds and these objectives, hence no propositions between our minds and the facts. Truth and falsehood are derivative properties of objectives, when these are considered as the objects of what we should now call propositional attitudes. Findlay says of this, 'Meinong's theory of truth is therefore a theory of *identity* or *coincidence.* The *same* objective which is factual ... reveals itself in a certain judgement or assumption ... ; the fact itself is true in so far as it is the object of a judgement'.[19] What is it, though, for an objective to be factual?

---

[15] Palmer [1988, *passim*] thinks it extremely serious. Sainsbury [1979, 20–25] suggests that, although it is a real difficulty for Russell, it just shows that he had a muddled conception of the construction of propositions.

[16] [Russell, 1903, §52].

[17] For example, [Baldwin, 1991; Cartwright, 1987; Dodd, 1995, §3; Hornsby, 1999; David, 2001, p. 684]. The now-common but then unknown vocabulary of truth-bearer and truthmaker has to be understood as involving truth-aptness rather than actual truth; a truth-bearer may be false.

[18] [Findlay, 1933, ch. X sec. IX].

[19] ibid. 88: italics in the original.

Factuality is, Meinong says, 'a fundamental property which admits of no definition'.[20] It could hardly be clearer how easy it is to move, without noticing, from the identity theory into primitivism.

But an identity theory of *truth* is unavailable to Moore and Russell, for they give an identity account of *all* propositions (truth-bearers), true and false, between which no distinction can be drawn merely by appeal to identity with some combination of propositional constituents (truthmakers); hence the need for a further property to accomplish the task. Russell himself appears to recognize the difference between primitivism and an identity theory in a slightly later presentation of his views in which primitivism has mutated from its earlier version, in which false propositions merely lack the property of truth, into one in which true propositions have one property and false another (though he gives no attempt to account for the opposition of truth and falsity).[21] In both versions, it is clear that the idea that truth is a primitive property is imposed by the failure of identity between representation and represented to provide a distinction between truth and falsehood.

1.3 Identity and coherence: Blanshard, Bradley, Joachim

The philosopher against whom both Moore and Russell took themselves to be principally reacting, and who exemplified in their minds the views they were rejecting as pernicious, was F. H. Bradley. Bradley is presented in the great majority of philosophy textbooks as a coherence theorist, and this is certainly how Russell understood him. He is also often implicitly presented as a metaphysician and no logician.[22] But he wrote a major work on the philosophy of logic, and his views are mostly ignored now because he, like Mill, attacked the idea that logic could be both formal and adequate to represent reasoning. In this major work (which incidentally convinced Russell that the logical form of universal propositions is hypothetical) he indicates that logic requires a correspondence theory.[23]

Despite this, he himself did not endorse correspondence, since in characteristic fashion he regarded logic as an inadequate key to metaphysics. His own account of truth is reached by a reaction against the correspondence theory (which he calls the 'copy' theory). There is a set of problems clustering around the notion of judgement which Bradley sums up succinctly just following his seeming endorsement of that theory:

> How then are ideas related to realities? They seemed the same, but they clearly are not so, and their difference threatens to become a discrepancy. A fact is individual, an idea is universal; a fact is substantial, an idea is adjectival; a fact is self-existent, an idea is symbolical. Is it

---

[20] Quoted by Findlay, ibid. 76.

[21] [Russell, 1904a, 473f].

[22] He is, for example, lampooned by Ayer in his assault on metaphysics [1936, 36], and ignored by the Kneales in their magisterial history of logic [Kneale, 1962].

[23] [Bradley, 1883, 41f].

> not then manifest that ideas are *not* joined in the way in which facts are? Nay the essence of an idea, the more it is considered, is seen more and more to diverge from reality. And we are confronted by the conclusion that, so far as anything is true, it is *not* fact, and, so far as it is fact, it can never be true.
>
> [Bradley, 1883, 43f]

The word 'fact' here indicates a truthmaker, and 'idea' a truth-bearer. The suggestion is that, because of the inherent limitations of symbolism, it is impossible ever to have a true judgement in the sense that it accurately reflects the reality with which it deals. The correspondence theory, applied to symbolic thought and taken quite literally, commits us to the view that no judgement is ever actually true.

This train of thought is behind the most important of a string of problems which Bradley marshals in his consideration of the correspondence theory.[24] First, judgements about the past and the future cannot be the result of copying.[25] Second, the very facts whose copying is supposed to give us truth are themselves 'the imaginary creatures of false theory', whose seemingly independent existence is merely the result of projecting on to the world the divisions imposed by thought, whereas if thought is to be capable of truth those divisions must exist independently of thought itself.[26] Third, '[d]isjunctive, negative and hypothetical judgements cannot be taken as all false, and yet cannot fairly be made to conform to our one type of truth', and neither can '[u]niversal and abstract truths'.[27]

Bradley then moves on to the pragmatic theory of truth, and suggests a fourth objection: that at bottom, both theories commit the error of defending the supposition of a 'truth which is external to knowledge' and a 'knowledge which is external to reality'.[28] The argument that this is an error seems to turn on the claim that the supposition involves a vicious circularity: for 'p' to be true, it must be true also that 'p' is a copy of p, or that believing 'p' is advantageous (and so *ad infinitum)*; and for 'p' to be known to be true, it must also be known that 'p' is a copy of p or that believing it is advantageous (and so *ad infinitum*).

The third of these objections was serious, and required the invention of the theory of truth-functions for a plausible answer to be provided. The first, though, might be easily dismissed as arising from a misunderstanding. But Bradley's discussion of the correspondence theory is infected by a strain of anti-realism, which would be expected from one who is committed to idealism. (This helps to explain why he calls it the copy theory, since he appears to assume that it is a theory of the genesis as well as of the nature of truth.) In his view, truth cannot be verification-transcendent, and, on the correspondence theory, must be obtained by a process of copying reality. But the correspondence theory is usually associated with metaphysical realism. This objection thus rests on a far wider disagreement.

---

[24] [Bradley, 1907].

[25] ibid. 107.

[26] ibid. 108.

[27] ibid. 109.

[28] ibid. 111.

We have already seen Moore making the claim on which the fourth objection rests. Philosophers divide over whether to take it seriously, either in general (for it threatens every attempt to give an account of what truth consists in) or in application to the correspondence theory.[29] But if we take the notions of correspondence and fact at face value, so that the fact is an independently existing counterpart of the proposition and there must be a further fact for each true proposition about correspondence to correspond to in order to be true, then the regress will be vicious. When one thinks of the elaborate apparatus employed in the *Tractatus Logico-Philosophicus* to avoid such a regress, in particular the distinction between showing and saying, one could reasonably wonder whether the correspondence theory's regress can be dismissed without following that book in making the theory unstatable.

The second objection is the most foreign to a modern reader, who may be surprised to learn that these grounds of Bradley's dissatisfaction with the correspondence theory were shared by that hero of modern logicians, Frege:

> A correspondence, moreover, can only be perfect if the corresponding things coincide and so are just not different things... It would only be possible to compare an idea with a thing if the thing were an idea too. And then, if the first did correspond perfectly with the second, they would coincide. But this is not at all what people intend when they define truth as the correspondence of an idea with something real. For in this case it is essential precisely that the reality shall be distinct from the idea. But then there can be no complete correspondence, no complete truth. So nothing at all would be true; for what is only half true is untrue. Truth does not admit of more and less.
>
> [Frege, 1918, 3]

Frege follows this initial argument against the correspondence theory with the consideration of an obvious response, that all that is required is 'correspondence in a certain respect' (loc. cit.). To this, where Bradley would have rejected such talk of 'respects' with his second objection, Frege provides Bradley's fourth objection, of vicious circularity, which he supposes gives us, when generalized, and buttressed by the assumption that truth is not a matter of degree, good reason not only to reject the correspondence theory but to maintain that truth is '*sui generis* and indefinable', a 'property of a thought',[30] thus apparently abandoning his earlier and more well known view that the True is an object named by sentences. Frege, then, like Moore, thinks that all attempts to define truth in these kinds of ways involve a vicious circle, and reverts to primitivism. But Bradley tried instead to bring thought and reality together, to do justice to the idea that when we think truly, *what we are thinking is what is the case*, and follows his rejection of correspondence by expounding his own largely unrecognized view of truth:

---

[29] Examples of both sides of the divide are provided by Ralph Walker. In his [1989, 99] he asserts without argument that the correspondence theory's regress is non-vicious. In his [2001, 150–1], he attempts to *show* that the regress is non-vicious for the correspondence theory while vicious for the coherence theory. The argument appears to turn on a confusion of the correspondence platitude (which all theorists of truth can agree to affirm) and the correspondence *theory*, which requires some serious account of the nature of the correspondence.

[30] [Frege, 1918, 4; 6].

> The division of reality from knowledge and of knowledge from truth must in any form be abandoned. And the only way of exit from the maze is to accept the remaining alternative. Our one hope lies in taking courage to embrace the result that reality is not outside truth. The identity of truth knowledge and reality, whatever difficulty that may bring, must be taken as necessary and fundamental.
>
> [Bradley, 1907, 112–3]

This view has come to be called the identity theory of truth. But what does this slogan 'the identity of truth and reality' amount to?[31] It looks as if Bradley means that truth consists in the identity of some x with a reality that thereby makes x true; and the identity theory's most general form is that we have truth iff the truth-bearer is identical with the appropriate truthmaker.

We can see that Bradley's identity theory of truth arose out of his second objection to the correspondence theory: that its view of facts is based on the illusion that a proposition can be true by corresponding to part of a situation even though it rips that part from its background and separates it up into further parts which are not separate in reality. Because he rejected this fragmented world in favour of a monistic ontology, not of a Parmenidean sort but one in which reality is itself a coherent whole of differentiated but not separate parts, his identity theory of truth allows him to employ a coherence theory of justification, and entails that coherence is a test of truth. In fact, in his view, it is *the* test.[32] The result is that much of his discussion of truth is conducted in coherentist terms (e.g. 'system'). It is easy to be misled; and when Russell launched his famous attack on the coherence theory,[33] while he focused ostensibly on Harold Joachim's version of it, it is clear that his real target was Bradley, whom he took to be an archetypal coherence theorist. (Russell's seminal role in the subsequent discussion explains the already-remarked mistakes of the textbooks.) Russell's focus on Joachim may not have been entirely misdirected, however, since Joachim's theory is obviously inspired by Bradley's, and he too features, along with Brand Blanshard, as a textbook coherence theorist — indeed, the theories of Joachim and Blanshard are so similar that we may usually treat them as one. And yet, like Bradley's, their theory has an important feature that is either overlooked or underemphasized in the textbook characterizations of coherence theories of truth. For although, in contrast to Bradley, they insist that the nature (and not just the test) of truth lies in coherence, they resemble him in claiming that the genuinely coherent system of belief will be identical with reality.

---

[31] The presence of the word 'knowledge' in Bradley's version of the slogan is to be explained through the already-noted verificationist element in his idealism. We ignore it here as an unnecessary hostage to fortune.

[32] [Bradley, 1909, 202].

[33] [Russell, 1907].

However, while Bradley's appeal to coherence follows from his identity theory, the direction of entailment seems to flow the other way for Joachim and Blanshard. They worked their way into a theory of truth through a theory of judgement that began with the idea that the nature of judgement needs to be understood teleologically – to understand judgement we need to understand the goal we have in making a judgement. They concluded that there are two such goals. One is, as Bradley insisted, that in judging we aim to judge that things are the way they are. The other, 'immanent', goal we have in making a judgement is to put an end to our inquiry. Of course, neither Blanshard nor Joachim thought that just any judgement would suffice for this. Only a properly justified judgement can satisfy our curiosity.[34] So we need to understand what it is for a belief to be justified in order to understand the nature of judgement.

For the early coherence theorists, a justified judgement is one that coheres with the rest of our beliefs. But what, exactly, is it for two things to 'cohere'? Do beliefs cohere as long as they are consistent? For both Joachim and Blanshard, a coherent set of beliefs must not only be consistent, but also form a unified and explanatory system.[35] Blanshard goes even further and claims that in the perfectly coherent system, that which lies at the end of all inquiry, each judgement will be necessitated by the others. The central feature in this account of justification, though, is its holism. For clearly no belief considered on its own could be coherent in this sense. Instead, it is a whole system of beliefs that is coherent and whether any particular belief is justified is a matter of whether the system of belief will be coherent, remain a *system*, after the inclusion of the belief. Thus for Joachim and Blanshard, as for Quine, the primary bearer of epistemic merit is a whole system of beliefs.

Moreover, this system of beliefs is a *teleological* whole because it has a unity that lies, in part, in the fact that its construction has the common goal of putting an end to the inquiry. As Joachim characterizes it, a teleological whole is '... a whole of parts such that each part contributes determinately to constitute the whole, and that the structural plan of the whole determines precisely the nature of the differences which are its parts...'[36] Given that the parts of the system are the beliefs that make it up, this conception of the system implies that the nature of a belief depends upon the whole of which it is a part. In fact, both Blanshard and Joachim insist that judgements do not have any determinate significance in isolation.[37]

---

[34] [Blanshard, 1939, vol. 1: 489; Joachim, 1906, ch. 3].

[35] [Blanshard, 1939, vol. 2: 264; Joachim, 1906, 73–8].

[36] [Joachim, 1906, 9f].

[37] [Blanshard, 1939, vol. 2, 266; Joachim, 1906, 73, 93]. Although their overall position combines epistemological holism with meaning holism, there is no argument provided (as far as

But what do these claims about justification and judgement have to do with truth? Blanshard argued that the conception of our belief system as a teleological whole united under the goal of ending inquiry was incompatible with a correspondence theory of truth. He maintained that to suppose truth is correspondence is to suppose that the goal of establishing the systematic coherence of our beliefs is different from that of apprehending the nature of reality. However, if this is the case, what reason do we have for supposing that by pursuing the goal of coherence we are creating a system of thought that corresponds to some external reality? And if there is no reason to think that we are getting at the truth by constructing such a system, why should we stick with this method of inquiry? As Blanshard thinks that this method of inquiry is part of the nature of thought itself he concludes that the correspondence theory's failure to justify this procedure (by failing to tie it to truth) will leave us forever out of touch with reality.[38] Furthermore, for the correspondence theorist to identify the two goals of judging, it would have to be possible to justify a belief by comparing it to some independent, unconceptualized fact. But Joachim and Blanshard's coherence theory of justification was based on the claim that we have no access to unconceptualized facts that on their own could justify some belief. So there is no way for us to justify our thoughts by comparing them with such facts.[39]

Indeed, Joachim and Blanshard agreed with Bradley's second objection to the correspondence theory, claiming that both reality and our conception of it are teleological wholes whose unity would be destroyed, and their nature falsified, by what he had called the 'vicious abstraction' that such a theory entails.[40] As applied to truth, the idea is that we do not speak the truth if we say less than the situation we are talking about would justify, just as we do not speak the truth if we say more, or something entirely different. Their hostility to any such abstraction ensures that, when their views are consistently carried through, at most one proposition can be true — that which encapsulates reality in its entirety.

The identity theory in this version has the advantage that it can meet a condition of any theory of truth, that it must make room for falsehood, the condition which diverted other potential identity theorists in the direction of primitivism; for it can account for falsehood as a falling short of this vast proposition and hence as an abstraction of part of reality from the whole. The result is that all three adopted the idea that there are degrees of truth: that proposition is the least true which is the most distant from the whole of reality.[41] Adopting this doctrine at least allows some sort of place for false propositions and the possibility of distinguishing worse

---

we are aware) from one form of holism to the other. They seem to draw their motivation for both positions from their conception of the goal of thought as the identification of thought with reality and the idea that this identification can occur only if the reality has been made intelligible.

[38] [Blanshard, 1939, vol. 2: 267ff].

[39] Joachim carefully spells this out in [Joachim, 1906, ch. 2]. The question remains, of course, as to whether the two goals of judging really do need to be identified as Joachim and Blanshard insist.

[40] [Joachim, 1906, 36ff; Blanshard, 1939, vol. 2: 266–7].

[41] [Bradley, 1909; Blanshard, 1939 vol. 2, ch. 27; Joachim, 1906, 85–121].

from better.[42] However, the consequence of this is that all ordinary propositions will turn out to be more or less infected by falsehood because they fail to reach this ideal of inclusiveness. It is also unclear how such a theory can distinguish between the degrees to which different beliefs are false and so explain how we can be led towards the truth. For his part, Joachim seems to claim that the ground of falsity and error in particular judgements lies precisely in a failure to see that all such judgements are only partial truths.[43] As Russell was quick to point out, this suggestion entails that if someone asserts that some birds have wings, while confidently believing that the assertion is true, the assertion must be false.[44]

Russell also objected to the coherence theory on the grounds that it would be easy to create coherent systems of propositions that contain falsehoods.[45] For example, the claim that no birds have wings undoubtedly belongs to some coherent system of propositions, though presumably one that is constituted by a vastly different range of propositions from those most of us accept. As we have seen, although this objection would have been partly directed at Bradley it has no force against his identity theory of truth. Yet its force against Joachim and Blanshard's theory is also questionable. For a start, all these philosophers insisted that the relevant set of propositions was that of those actually believed.[46] And they insisted further that not just any set of noncontradictory beliefs counts as coherent. In particular, these coherent systems were called 'self-fulfilling' in part because the standards that a belief must reach to become part of the system were themselves part of the system. These standards could (and do) evolve over time.[47] So judgements that cohere must not only be justified (rather than merely adopted on a whim) but the standards of justification themselves get stronger and our inquiries more focused and effective as they progress and we learn more about the domain in question. Thus it is at least not as easy as Russell suggested to create a coherent set of beliefs that contain what we would consider falsehoods. Perhaps it is possible to do so if we start our hypothetical set with different standards for entry into the set than those we actually have. For Joachim and Blanshard, however, it is not possible to swap our standards of justification: they are part of the nature of thought.

Regardless of whether any or all of Bradley, Joachim and Blanshard can provide an account of falsehood and avoid the Bishop Stubbs objection, the metaphysical price of their theories is obviously high. The price has several components. One is this: it is clear that the idealist metaphysics built into Joachim and Blanshard's peculiar teleological account of judgement is what ensures that their claims about the identity of truth-bearer and truthmaker are at least not non-starters. Yet this metaphysical position on its own is too costly for most philosophers.

---

[42] Although philosophers have tended to share Russell's scorn for this idea (see, e.g., the Fregean argument quoted near the start of this section), variants of it keep turning up. See §4 below.

[43] [Joachim, 1906, 162].

[44] [Russell, 1907, 135].

[45] Because Russell used the example of a coherent system that contained the claim that the (in fact eminently respectable) Bishop Stubbs was hanged for murder, this objection has become known as the Bishop Stubbs objection. Russell [1907, 135–9].

[46] Lengthy investigations of whether this move protects the coherence theory from the Bishop Stubbs objection can be found in [Walker, 1989; Wright, 1995].

[47] [Blanshard, 1939, vol. 1: 490; Joachim, 1906, 76f].

Further, Bradley worried that the theory's sole all-describing proposition will still be infected by falsehood. For the nature of symbolism demands that it display reality's connected aspects by means of separate fragments, and it will itself both have to be, as an existent, part of reality and yet, as reality's description, separate from it. The only resolution of these difficulties which he could see was to go further in the same direction, concluding that the total proposition, to attain complete truth, would have to cease to be a proposition and *become* the reality it is meant to be *about*. This seems to be what Blanshard had in mind when he claimed that truth is 'thought on its way home'.[48] While Joachim and Blanshard agreed to such talk of identity, though, it was not meant to *replace* the claim that the nature of truth is coherence. They insisted that theirs is a coherence theory. It is just that the ideally coherent system of judgements will be identical with reality.

But Bradley drew more extravagant consequences from their shared hostility to abstraction. While this hostility initially motivates the identity theory, when it is allowed to remain unbridled, its implications begin to threaten even the theory itself. For although Bradley, Joachim and Blanshard all described their view of truth in terms of 'identity', and so justify talk of an identity theory of truth, this title is ultimately misleading in application to Bradley, since his theory ends up as eliminativist: on his anti-Hegelian view, reality transcends the rational, and turns out not to have a fact-like structure expressible in any propositional form at all,[49] so that when full truth is attained, the point of inexpressibility is reached. Hence Bradley, despite using the word 'identity' to describe his view, says as well that 'in the proper sense of thought, thought and fact are not the same' and talked of the attainment of complete truth in terms of thought's 'happy suicide'.[50] In effect, as the proposition approaches complete truth, it disappears altogether in favour of reality:

> But if truth and fact are to be one, then in some such way thought must reach its consummation. But in that consummation thought has certainly been so transformed, that to go on calling it thought seems indefensible.
>
> [Bradley, 1893, 152]

Bradley's metaphysical theory of truth, when its consequences are fully explored, thus turns out to be self-destructive. His metaphysics is such that he did not regard this as an objection. But it is likely to seem so to those — surely the overwhelming majority — unwilling to share his entire metaphysical vision, with the result that the theory appears to be merely an historical curiosity. However, we shall see that the fundamental ideal of the identity theory — securing truth by closing the gap between mind and world — has been recently revived and is once more influencing discussion.

---

[48] [Blanshard, 1939, vol. 2, 264].

[49] [Bradley, 1883, 590f].

[50] [Bradley, 1893, 150, 152].

1.4 Correspondence: Russell, Wittgenstein

Russell's attack on Joachim signalled a move away from the binary theory of judgement and its required primitivist account of truth. As early as 1904, he was articulating worries about primitivism in his long consideration of Meinong:

> It may be said — and this is, I believe, the correct view — that there is no problem at all in truth and falsehood; that some propositions are true and some false, just as some roses are red and some white; . . . . But this theory *seems* to leave our preference for truth a mere unaccountable prejudice, and in no way to answer to the feeling of truth and falsehood . . . .
>
> The fundamental objection may be simply expressed by saying that true propositions express *fact*, while false ones do not. This at once raises the problem: What is a fact? And the difficulty of this problem lies in this, that a fact appears to be merely a true proposition, so that what seemed a significant assertion becomes a tautology.
>
> [Russell, 1904a, 473]

After this succinct discussion of issues still surrounding the notion of truth, Russell reassures himself that primitivism is all right — 'What is truth, and what falsehood, we must merely apprehend, for both seem incapable of analysis' — and it turns out that 'our preference for truth' (which we have since learned to re-label as the claim that truth is a 'norm of assertion') is not 'a mere unaccountable prejudice' but is justified by 'an ultimate ethical proposition'. Still, it is clear that he is uneasy. By 1907 Russell's discomfort with the primitivism imposed on him by the binary theory was great enough for him to end his critique of Joachim by contemplating replacement of the latter with the multiple relation theory of judgement; and in 1910 he committed himself to the change.[51]

The new theory is developed against the background of his criticism of Meinong's primitivist account of truth and falsehood as properties of objectives. Falsehood, he now thinks, is the work of the mind and not an independent property: it is impossible to believe in the existence of real mind-independent objectives where a judgement is false, and this provides sufficient reason for not believing in them even where the judgement is true; furthermore, primitivism renders the true/false distinction a 'mystery'. In consequence, he opts for a new theory in which judgement is not binary but 'a multiple relation of the mind to the various other terms with which the judgement is concerned'.[52] It is clear that most of his dissatisfaction is not in fact with primitivism *per se*. It is rather with the idea that the world contains mind-independent falsehoods; and this is a consequence of the combination of the unity of the proposition with the doctrine of real propositional constituents, not of primitivism. Be that as it may, however, propositions, as the truth-bearing unified entities which figured in the binary theory, have disappeared altogether (although the vocabulary lingers); they have been displaced by propositional acts. A full account of the theory, together with its version of the correspondence theory of truth, is given in the last couple of pages of the paper. Using some of Russell's own words, it may be summarized thus: When we judge that, say, A loves B, we have 'before the mind' the person A, the person B, and the relation of loving, in such a way that the relation is not present 'abstractly' but as proceeding from A to B. The judgement is true when there is a corresponding complex object, A's loving B, and false when there is not. Russell thus endorses a correspondence theory of truth, in which the complex object (which Russell was soon to call a 'fact') to which a true judgement corresponds is something the theory presents as quite independent of that judgement itself.

---

[51] [Russell, 1910b].

[52] ibid. 122.

This presentation of the multiple relation theory embroiled Russell in a confusion of the problem of direction (how do we ensure that non-symmetrical relations like 'loves' go in the right direction?) with the problem of unity (how do we ensure that we have a proposition, and not a mere collocation of its individual constituents?), a confusion which was partly responsible for a rapid succession of different versions of the multiple relation theory, whose differences may be ignored here,[53] for in all its versions the theory is still dogged with the problem which had forced Russell to adopt primitivism as the suitable theory of truth for the binary relation theory of judgement.

The multiple relation theory was meant to circumvent the binary theory's problematic requirement of the mind-independent existence of all propositions constructed from real constituents. But once primitivism's apparatus for making the true/false distinction is no longer available, false judgement is rendered impossible, even on the multiple relation theory. That theory is made necessary by Russell's lingering attachment to the doctrine of real propositional constituents and the idea that a truthmaker is a set of objects unified by a relation which is itself one of those objects. But these views, combined with Russell's recognition of the need to distinguish between a judgement that A loves B and the mind's merely being simultaneously acquainted with A and love and B, are bound to lead to a collapse in the ability to employ the true/false distinction. To see this, suppose that it is true that aRb, i.e. that this unified 'complex object' exists. Now suppose that someone S judges that aRb. This judgement consists in the unification of S, a, R and b by the judging relation (call it 'J'). But on this account, all such judgements will be false owing to a failure of correspondence, since in the fact a and b are related by R, whereas in the judgement — according to an essential component of the multiple relation theory — a and b are not related by R (but by J). A natural response to this objection is to say that the multiple relation theory should be modified so that a and b are related by R inside the judgement, thus enabling correspondence to hold. But this modification has the consequence that no judgement can be false, since any judgement will unify its components and create the fact which makes it true. That is, either the judging does not include a suitable correspondent for the judged, in which case nothing can be true; or it does, in which case nothing can be false.

---

[53] They are discussed in detail in [Candlish, 1996], from which the current discussion has been condensed and modified.

A possible defence of Russell's combination of views at this point would be to say that all that this objection reveals is a serious unclarity in the notion of correspondence, and that Russell should have gone on to show how a non-unified collection of objects can correspond in the required sense to a complex object whose components are those objects. Even if this line is taken, however, the multiple relation theory still relies upon a mysterious power of the mind to assemble and arrange real objects (and modern versions of the theory which employ sequences to do the same job are no more than stipulations which merely disguise the mystery). Furthermore, it does nothing to address the difficulty concerning unity which had partly prompted the replacement of the binary by the multiple relation theory, that is, Russell's treatment of relations as substance-like objects and his consequent requirement that a relation play the inconsistent role of both the unifier and the unified: in the new theory this role has merely been transferred from the judged relation R to the judging relation J.

In 1918, Russell himself effectively conceded the first of these criticisms and admitted the problem of falsehood had not been solved; in 1924 he conceded the second.[54] In fact, within the metaphysics of logical atomism, it is vital for the possibility of false judgement, and indeed of a correspondence theory of truth, that there be some distinction between the real objects about which some judgement is made and the constituents of the judgement. It is thus hardly surprising that, soon after this concession, Russell abandoned both the doctrine of real propositional constituents and the multiple relation theory of judgement in favour of a mentalistic view of the nature of propositional constituents while retaining successive variants of the correspondence theory of truth.[55]

A feature of both the multiple relation and the correspondence theories is that Russell originally explained them only for non-quantified propositions; he abandoned the former before attempting its extension to include quantification, but it is clear that any such extension could not be straightforward, and poses a problem for the associated extension of the latter, since the nature of correspondence as originally explained would not apply where quantifiers are involved.[56] At first this seemed to Russell an advantage, since it opened the possibility of explaining one of the distinctions required by type theory (and at one stage he tried unsuccessfully to use it to solve the Liar Paradox). But it also posed a problem he never solved: his logic contained unrestricted generalizations about propositions,[57] and these require an infinite realm of propositions which exist independently of their contingent formulation by finite minds. A natural solution to the difficulty would be to say that the quantification extends over possibilities; but Russell himself made this solution unavailable by his repeated insistence that possibility is not fundamental but must be accounted for in terms of actuality.[58]

---

[54] [Russell, 1918, 198f; Russell, 1924, 170–3]. The latter concession is disingenuously presented as not a change but a clarification.

[55] [Russell, 1919; 1959, ch. xv].

[56] He eventually accommodated the latter by the postulation of 'general facts', in addition to singular facts, to be the truthmakers for quantified propositions; it is still far from obvious how this postulation is consistent with the doctrine of real propositional constituents. See [Russell, 1918, 206f].

In 1913 Wittgenstein had severely criticized the latest version of Russell's theory of judgement. While there is still argument about the nature of the criticism, there is no argument about the fact that Wittgenstein's views in the *Tractatus* were often formed in reaction to Russell's. He follows Russell in adopting a correspondence theory: but his accounts of propositions and truth are distinguished by their embodying ingenious treatments of the problems with which Russell had been struggling unsuccessfully.

First, Wittgenstein rejected the doctrine of real propositional constituents; instead, propositional constituents are Names, not Objects. (The expressions are capitalized in this paragraph to draw attention to the fact that these are no ordinary names or objects, but the end-points of analysis required by a theory.) This immediately relieved him of the worry that the demand that the proposition be a unity would make every proposition true, the worry which had driven Russell, first, into primitivism, and then, following his rejection of primitivism, into the multiple relation theory of judgement. Wittgenstein could allow propositions to be unified without risking the creation of their truthmaking facts, whose constituents are Objects, not Names.

Secondly, Wittgenstein had a coherent account of propositional unity itself (and, incidentally, of the corresponding unity of the truthmaking fact). To take Russell's example, suppose that it is true that A differs from B. Russell had understood this proposition as consisting of three things, A and difference and B; as we saw, he was unable to explain how these saturated objects could form either a proposition, or, after the abandonment of the binary theory, a fact. (Bradley had seen the difficulty, arguing that if one thinks, as Russell does, of relations as a sort of object, the demand for propositional unity sets off a vicious infinite regress as one endeavours vainly to find a relation which will not itself need further relating to its relata.) Part of Wittgenstein's strategy is obvious enough: it was to reject Russell's treatment of relations as saturated objects. But he followed Russell in rejecting too Frege's idea that relations are unsaturated constituents of propositions. This left the problem — a solution to which is vital for a defence of a correspondence theory of truth — of accounting for the truth of relational statements.[59] He unravelled this tangle by maintaining that both facts and propositions are unified by relations which do not figure in them as constituents.

---

[57] Compare the point about the axiom of infinity in §1.2.

[58] The matters touched on in this paragraph are well explained in [Hylton, 1990, 355f].

[59] As we saw, for Russell these were *all* statements, but in fact Wittgenstein's solution, while compatible with this Russellian view, does not require it, since, if it works at all, it applies just as well to monadic propositions treated as fundamentally so.

The crucial remarks in the *Tractatus* for understanding Wittgenstein's views on these matters are these:

> The propositional sign consists in the fact that its elements, the words, are combined in it in a definite way.
>
> The propositional sign is a fact.
>
> [Wittgenstein, 1921, 3.14]; Ogden's translation)

> Not "The complex sign 'aRb' says that a stands to b in the relation R", but rather, that "a" stands to "b" in a certain relation says that aRb.
>
> (ibid. 3.1432; SC's translation.)

In the *Tractatus*, then, propositions contain no names of relations, and in particular the relation which unifies them does not appear as a propositional constituent but is exhibited by the relation of the names to each other. Correspondingly, in the truthmaking fact, relations are not a further kind of object demanding to be related to the other constituents of the fact. The symbolizing of a relational fact is accomplished by the construction of another relational fact which is isomorphic in its structure. Propositions, on this account, are a kind of picture.

Along with this solution of the problem of unity we get a definite account of the nature of the correspondence involved in truth: a proposition is true just if the arrangement of its constituent names is isomorphic to an actual arrangement of objects, with a 1:1 relation of names to objects; it is false when the arrangement is merely possible but not actual. The pictorial relation is not itself stated, but shown; in Wittgenstein's view it cannot be stated, since it involves logical form, which is presupposed in any proposition at all. In this way, Wittgenstein was able both to maintain a correspondence theory and to evade the Bradley/Frege argument that the theory involves a vicious infinite regress: the evasion is accomplished by making the theory unstatable.

Of course, propositions expressed in natural language look nothing like pictures. Wittgenstein dealt with this understandable reaction in three stages. First, the notion of *picture* is generalized to embrace propositions.[60] Second, the claim that the proposition is a picture is restricted to the *fully analysed* proposition; when a proposition's deep form is revealed, it will be shown to consist of nothing but Names of Objects. The analysis is accomplished by repeated application of Russell's Theory of Definite Descriptions to everyday names until the real ones are reached. Third, the pictorial account is restricted to atomic propositions; molecular propositions are truth-functions of atomic propositions. This meets one of Bradley's objections to the correspondence theory: the theory can be restricted to a base class of atomic propositions from which all others can be generated.

---

[60] This is accomplished in §§2.1–2.19; it is well described in Pears 1977.

The pictorial account also provided Wittgenstein with a way of dealing with another problem which had been puzzling Russell, namely, why belief contexts are not truth-functional despite their appearing to contain propositions. Russell had concluded that, not being truth-functions of the propositions believed, belief statements reveal a new kind of atomic fact.[61] Wittgenstein, rejecting both the binary and the multiple relation view of judgements, concluded that they are neither atomic nor molecular;[62] rather, they are not propositions at all, but disguised deployments of the pictorial relation — the real form of 'A believes that *p*', is '"*p*" says that *p*',[63] which on the *Tractatus* account is not a proposition at all but just an attempt to say what can only be shown.

In this way Wittgenstein deals all at once with a great range of problems surrounding the notion of truth. But the metaphysical price paid is very high, and it was not long before even he decided that he was not prepared to pay it.

1.5 Pragmatism: Dewey, James, Peirce

At the same time that Russell and Bradley were arguing with each other about truth, both were arguing with William James. Like other pragmatists, James rejected the idea that there are fixed, ideal structures of thought such as we saw (in §1.3) figuring in the theories of Blanshard and Joachim. Moreover, the pragmatists denied that thought and reality are such that they could be, even if only ideally, identical (as Bradley as well as Blanshard and Joachim seemed to maintain). On the other hand, like these theorists, the pragmatists attacked as false abstractions the correspondence theory's twin notions of truth and facts as external to justification. Yet despite their combative stance, they shared the common approach of devising a theory of truth on the basis of a theory of judgement.

The American pragmatists were largely united under a maxim expressed by Peirce:

> Consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object.
>
> [Peirce, 1878, 258]

This maxim served as a way to 'make our ideas clear' and to brush aside metaphysical games of make-believe and philosophical arguments whose resolution could have no practical significance to our lives. Applied to judgements, it obviously implies that they are to be individuated according to their practical causes and effects. Yet although the pragmatist theory of judgement originates with Peirce's maxim, there is a significant difference between his pragmatic theory of truth and the sort of conception of truth found in the works of William James and John Dewey.[64]

---

[61] [Russell, 1918, 199].

[62] The rejections of the binary and multiple relation theories are at [Wittgenstein, 1921, §§5.541–2 and §5.5422 respectively].

[63] ibid. §5.542.

Peirce applied the pragmatic maxim directly to our concept of truth and argued that the only experiential and pragmatic concepts we have to guide us here are the notions of doubt and belief:

> But if by truth and falsity you mean something not definable in terms of doubt and belief in any way, then you are talking of entities of whose existence you can know nothing, and which Ockham's razor would clean shave off. Your problems would be greatly simplified, if, instead of saying that you want to know the "Truth", you were to say that you want to attain a state of belief unassailable by doubt.
>
> [Peirce, 1905, 279]

Thus, truth is the property of those beliefs that are unassailable by doubt and that therefore register the fact that we have formed a settled opinion. The true beliefs are those that are held at the end of our inquiry. Like Joachim and Blanshard, Peirce emphasizes that the immanent goal of inquiry is the suspension of doubt and identifies the beliefs reached at the end of an ideal inquiry with the set of *true* beliefs.

However, it would be a mistake to see Peirce's theory of truth as a pragmatically construed coherence/identity theory. Peirce did not think that our final set of beliefs would be identical with reality. Instead, he held the optimistic belief that only a properly scientific inquiry would successfully create a stable end to doubt and inquiry and at the same time reveal the nature of reality. Admittedly, Peirce gave 'reality' a revised definition. 'The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real. That is the way I would explain reality.'[65] It is important to see that although Peirce shares the coherence theorist's intuition that reality is just what is 'represented' by the set of beliefs at the end of inquiry, his motivations are entirely different. For Peirce, the pragmatic maxim about meaning dictates that 'reality' be construed so that the real is, ideally, attainable within the realm of experience. There is no room for a reality that is in principle epistemically unreachable and plays no role in guiding our inquiry. By redefining 'reality' in this way, Peirce has reinterpreted the idea of true beliefs corresponding to reality. Because reality is what the ultimate set of beliefs will say exists, we have been given a guarantee that the true beliefs are those that correspond to reality.

---

[64] James and Dewey themselves had slightly different theories of truth; in particular, Dewey seems more willing to straightforwardly equate truth with verification (see [Dewey, 1948, 159f]). However, given the restrictions of space, we have focused on the claims to which they both would have agreed.

[65] Peirce 1878: 268.

Nevertheless, it was these close ties to the coherence theory that led James and Dewey to develop the pragmatist line further and so abandon Peirce's theory of truth. Both denied that there would be some ultimate or final point of view that would contain all and only the truths:

> But owing to the fact that all experience is a process, no point of view can ever be the last one. Every one is insufficient and off its balance, and responsible to later points of view than itself.
>
> [James, 1904, 55]

In effect, this point from James amounts to a criticism of Peirce's conception of inquiry from within pragmatism itself. How are we to understand the notion of an ideal end to inquiry in pragmatic terms? How are we to tell that we have reached the end rather than merely fooled ourselves into thinking that we have because the game has started to get boring and difficult?[66]

James and Dewey both considered the essential feature of their pragmatism to be their conception of judgements as tools of our own making that are designed to help us cope with our surroundings.[67] On this view, we should not expect some endpoint to inquiry. What counts as useful is, of course, interest-relative and we have no reason to suspect that we have some fixed set of interests that will determine some final, ultimately useful system of beliefs. So, as our interests endlessly change, new needs and questions will arise and the process of inquiry will continue on.

This view of the goal of judging leads quite naturally to a theory of truth. If the goal of judging is to help us cope, then the true judgements are those that succeed in this goal. James famously defined truth on just these lines.

> 'The true', to put it very briefly, is only the expedient in the way of our thinking, just as 'the right' is only the expedient in the way of our behaving. Expedient in almost any fashion; and expedient in the long run and on the whole of course ... .
>
> [James, 1907b, 222]

Yet, such a theory of truth faces some obvious criticisms and, again, Russell was one of the first to raise them. Russell complained that James's theory entailed that we could not work out whether a belief is true until we worked out whether it would be useful to believe it.[68] But James claimed that their theory was not one of justification, not meant to provide us with a *criterion* for deciding which of our beliefs are true.[69] Even so, Dewey maintained that some belief may be true now, even if we do not now know whether it will turn out to be useful or not. It still has the ability to work now and this will come to light as it is tested and relied on.[70] To the extent the pragmatist theory does provide criteria for determining whether a belief is true or not, the criteria would be the same as any of us would endorse; namely weighing evidence, checking for consistency, inspecting the world and so on.

---

[66] [Rorty, 1986, 339].

[67] For an example of the claim that the pragmatists' theory of judgement is prior to their theory of truth see [Dewey, 1910, 165].

[68] [Russell, 1908, 135f].

[69] [James, 1908, 106f].

James gives us another definition of truth to help us see this. What, James asks, is the 'cash-value in experiential terms' of the notion of a true judgement? 'The moment pragmatism asks this question, it sees the answer: *True ideas are those that we can assimilate, validate, corroborate and verify. False ideas are those that we can not.*'[71] The problem with this pragmatist response is that their position now seems to be pulled in a number of different directions. On the one hand, the true beliefs are those that are expedient. On the other hand, the true beliefs are those that are or will be verified. Moreover, the pragmatist faces yet another obvious criticism, again voiced by Russell. Russell also complained that most of us do not want to know whether it is useful to believe that God exists, but whether God really exists.[72] In other words, the common understanding of truth, claims Russell, is one in which the belief that God exists is true if and only if God exists. How is the pragmatist going to account for this while maintaining both that the true is the useful *and* that the true is the verified or verifiable?

To answer this question one must realize that the pragmatists had a much broader sense of utility in mind than might be at first suggested by their definition of truth as expediency. For a belief to pay or be useful, it must cohere with our other beliefs and the beliefs of others, it must enable us to cope with the objects the belief is about and it must not lead to perceptual expectations which have been disappointed. In other words, coping is precisely a matter of unifying and explaining our experiences and other beliefs. Moreover, the content of a belief is to be understood pragmatically and so the meaning of a belief just is the perceptual expectations it creates, the actions it disposes us to perform and the inferential relations it has to other beliefs. Construing belief content in this way means that verification, assimilation, corroboration and the rest are signs that the belief fits into the network of belief, perception and action and helps grease the mechanisms that make our daily living possible.[73] Thus, beliefs which meet all of these constraints are satisfying to us and so, in a sense, give us what we want.[74]

Nevertheless, postulating truth as verification and utility seems to leave truth floating blissfully free from reality. If truth is as the pragmatist says it is, why should we think true beliefs let us in on how the world really is? For a start, James and Dewey agree that 'getting reality right' is an essential part of being true.

---

[70] [Dewey, 1910, 163].

[71] [James, 1907b, 212f].

[72] [Russell, 1908, 143].

[73] See, for example, [Dewey, 1948, 157].

[74] The notion of utility implied here was often emphasized by the pragmatists and puts paid to a common objection that goes back to G. E. Moore [1907], who criticized the obviously dubious claim that all and only the true beliefs are useful in getting us what we want. (There are many cases where it is more useful, in a *narrow* sense, to have a false belief than a true one.)

> If the reality assumed were cancelled from the pragmatist's universe of discourse, he would straightaway give the name of falsehoods to the beliefs remaining, in spite of all their satisfactoriness. For him, as for his critic, there can be no truth if there is nothing to be true about.
>
> [James, 1908, 106]

> I am of course, postulating here a standing reality independent of the idea that knows it. I am also postulating that satisfactions grow *pari passu* with our approximation to such reality.
>
> [James, 1907a,: 88]

Reality is independent of experience in that it does not require a belief in it in order to exist. In fact the pragmatist insists that the existence of certain situations, such as that the tree in front of me is a eucalypt, is really the only explanation of why it is useful to hold the belief that the tree in front of me is a eucalypt.[75] So, in a sense truth is a correspondence relation between beliefs and reality. As Dewey argues, the pragmatist merely explains in more concrete detail than is usual what the nature of this correspondence is. (The Procrustean taxonomies of the textbooks convert the overlapping views of real philosophers into artificially sharp contrasts.) This correspondence relation is constituted by the relations of verification and coping that hold between the belief and reality.[76] It is the whole process of being verified and validated and generally of 'agreeing' with reality.

Yet, both Dewey and James insist that reality cannot go beyond experience. The real is that which is or can possibly be experienced. This is why the first of the two quotes in the previous paragraph has the proviso 'from the pragmatist's universe of discourse'. Pragmatists do not believe in facts that stand beyond all experience and that make our beliefs true. They do, however, believe in independent facts within experience such that when this reality 'comes, truth comes, and when it goes, truth goes with it'.[77] With reality conceived like this, it makes no sense to wonder, when we have done all we can to verify, assimilate and corroborate some belief, whether this belief may still be false. As Peirce would say, to speculate so would be to engage in a typical philosopher's make-believe. While it is easy to sympathize with this hostility to speculation, one can see that it risks, despite the pragmatists' express desires, committing them to idealism. However, perhaps the objection to pragmatism which lingers longest is the sense that, for all their protestations to the contrary, the pragmatists at bottom identify truth with what is convenient. While there is clearly a nugget of insight in the pragmatist account, what is omitted is a clear view of what it is which makes truth a useful property of beliefs without being mere usefulness itself, and which also gives a point to the concept, a point gestured at by later philosophers' talk of truth's being a normative end of assertion. This clear view begins to emerge with the work of Frank Ramsey.

---

[75] [James, 1909, 8].

[76] [Dewey, 1910, 158f].

[77] [James, 1908, 106].

1.6 Redundancy: Ramsey

In his analysis of judgement, F. P. Ramsey agreed with Russell that a judgement must involve the mind's being multiply related to a number of objects. However, he also agreed with Wittgenstein that 'A judges that *p*' is really of the form "'*p*" says *p*'. This suggested to Ramsey that the whole problem of judgement really reduces to the question 'What is it for a proposition token to have a certain sense?'[78] Moreover, he realized that neither Russell nor Wittgenstein had given an answer to this question.[79] Like the pragmatists, Ramsey concluded that it is the behavioural causes and effects of holding a certain belief that constitute the fact that it has a certain content. Yet his conclusion was not identical to theirs, for he seems to have been the first to suggest a version of what has since been labelled 'success semantics': the view that a belief has the content that *p* iff *p*'s obtaining would result in the success of the actions we perform on the basis of that belief (together with some desire).[80] That is, truth is the property of (full) beliefs such that if all the beliefs which combine with any desire to cause an action have it, then that action will succeed in achieving the object of that desire. *This* is the kernel of truth in pragmatism. And, importantly, Ramsey thought that only this theory of content could explain why it is that we want true beliefs — namely, because true beliefs, more often than false ones, lead to the satisfaction of our desires.[81] The pragmatists, on the other hand, removed the need to explain why it is that true beliefs are more likely to get us what we want, by combining their theory of content with an identification of truth with utility. Ramsey, however, took a strikingly different approach to the theory of truth.

Ramsey argued that once we have an analysis of judgement there is no further problem of truth to be solved. However, this is not because he thought, like almost all the philosophers we have examined so far, that the correct theory of truth follows from the correct theory of judgement. Rather, Ramsey seems to have been the first to suggest that the theory of truth can be given independently of the theory of judgement.

Normally, he claimed, the bearers of truth and falsity are taken to be propositions. If we focus on propositions, assuming for the moment that we have solved the question as to why certain sentences express certain propositions and why certain mental acts are beliefs that p, then we have two contexts of truth predication to consider. In the first case we know exactly what proposition is being called true. About this case, Ramsey says, 'It is evident that "It is true that Caesar was murdered" means no more than that Caesar was murdered, and "It is false that Caesar was murdered" means that Caesar was not murdered.'[82] Accordingly, he believed that in this context both 'true' and 'false' were redundant predicates. We express the same content when we assert that a proposition is true as we do when we just assert the proposition itself (and *mutatis mutandis* for attributions of falsity).

---

[78] [Ramsey, 1923, 275].

[79] Russell's failure is noted on p. 142 of [Ramsey, 1927]; Wittgenstein's theory is examined in [Ramsey, 1923, 274–9].

[80] [Ramsey, 1927, 143f]. For a recent version of success semantics, see [Whyte, 1990]. For criticism and response, see [Teichmann, 1992; Whyte, 1992; Brandom, 1994; Whyte, 1997; Dokic and Engel, 2002; Daly, 2003; Mellor, 2003, 217–20].

[81] [Ramsey, 1927, 148].

In the second sort of context, 'true' is not eliminable in the same way. Consider a case where we say 'Everything Newton says is true'. In this case we mean something like 'For all p, if Newton says p, then p is true' (where the variable 'p' ranges over propositions). If we were to straightforwardly eliminate 'is true' from this sentence, the result, on its most obvious interpretation, would be ungrammatical. In the sentence 'For all p, if Newton says p, then p', the variable 'p' in its final occurrence is occupying a position in which only the substitution of a sentence would make sense. But if the quantifier is given its standard objectual reading, then a name is required in that position, and the whole sentence makes no more sense than the sentence 'If Newton said that snow is white, then Bob'. Nevertheless, Ramsey thought that even in this case we can assert something with the same content as 'Everything Newton says is true' without using a truth-predicate. For example, if we restrict ourselves to propositions of the form aRb, 'then "He is always right" could be expressed by "For all a, R, b, if he asserts aRb, then aRb", to which "is true" would be an obviously superfluous addition.'[83] Ramsey thought that the sentence he proposed overcame the problems faced by the approach that merely deleted 'is true' because his candidate sentence allows us to use the verb within the sentence itself, so that the consequent of the conditional is a sentence that is used rather than mentioned. Thus, he concluded, we need to discover (through the analysis of judgement) all the different forms propositions take, so that for each form we can construct a universal statement like that offered for aRb. Having done this we could conjoin all these universal statements, thus capturing the content of 'Everything Newton says is true'.

But if we understand quantification in the usual objectual way then Ramsey's paraphrase does not work. For, on this understanding, when we quantify using the variables a, R and b we are quantifying into name-position.[84] So even if the first occurrence of 'aRb' in the sentence Ramsey proposes is in name-position (so that if we replaced the variable with a name it would be a name for a propositional form[85]), the second occurrence must be read the same way and so, again, the sentence is ungrammatical. Alternatively, if we treat each of a, R and b as occupying name-positions, then replacements for 'aRb' would be collections of names and not propositions. Perhaps even more worrying is that Ramsey's proposed paraphrase requires us to list an infinite number of propositional forms.

---

[82] ibid. 142. Frege had made the same point about sentences some years earlier: "It is also worth noticing that the sentence 'I smell the scent of violets' has just the same content as the sentence 'It is true that I smell the scent of violets'." [Frege, 1918, 6]).

[83] ibid. 143.

[84] We follow Ramsey in treating these as variables.

[85] We shall ignore some notorious problems concerning form. For Russell's difficulties, see [Pears, 1977; Hylton, 1990, 344f; Candlish, 1996, 118–24]. For more general treatments, see

While these problems may seem to undermine Ramsey's redundancy account of truth, they can be avoided if we can find a way to read the quantifiers and their bound variables that makes sense of statements like 'For all p, if Newton says that p, then p'. Later, Ramsey himself suggested that statements with variables in sentence positions can be grammatical. In fact, he used this idea to put forward a simple definition of truth:

> A belief is true iff it is a belief that p and p.

Ramsey claimed that although this definition looks ungrammatical any instance of it will be grammatical because the sentence we substitute for 'p' will always have a verb within it, so that, again, the consequent of the conditional is a sentence that is used rather than mentioned.[86] If this is right, and we can make sense of Ramsey's propositional variables, then we would also have an account of the troublesome generalization contexts like 'Everything Newton says is true'. To elucidate these variables, Ramsey proposed that we see them as akin to pronouns, except that, instead of occurring in name position, they occur in sentence position. Remarkably, given the subsequent development of this idea (see §3.3), he made the explicit terminological suggestion that we think of them as 'pro-sentences'.[87] He also offered an explanation of why we cannot eliminate the truth-predicate from the troublesome generalization contexts. Ordinary language treats propositional variables as pronouns and not prosentences, so where propositional variables would occur unadorned in logical notation, in ordinary language we need to use the truth-predicate to turn a name into a grammatical sentence. So even though there is a definition of the truth-predicate, the peculiarities of English prevent us from eliminating 'true' from all the contexts in which it occurs.

Ramsey's account is breathtakingly elegant and highly suggestive. In particular, his idea that 'The proposition that p is true' and 'p' are in some strong sense equivalent seems to identify an important fact about truth. Further, if the sketch he offered for a way of understanding propositional variables could be made out in more detail, then a simple definition of the truth-predicate would be available. It could only be a matter of time before his ideas were taken up.

[Smiley, 1983; Oliver, 1999].

[86] [Ramsey, 1927–9, 9-10].

[87] ibid. 10.

2 MIDDLE VIEWS

2.1 Introductory comments

Ramsey's prescience in suggesting a redundancy theory, and separating the question of the semantics of belief from that of the nature of truth, is all the more impressive for coming at a time when the theory of judgement was almost universally allowed to dictate the terms on which a theory of truth could be constructed. Despite a change of vocabulary, this underlying approach remained the norm even through the change in methodology witnessed in the middle years of the century and later labelled 'the linguistic turn'. In the theory of truth this change meant that the focus of debate shifted to questions surrounding the use and meaning of the truth-predicate. However, despite a common linguistic focus, there was little consensus during this period about what the purpose of a theory of truth is. In particular, two important dimensions of disagreement concerning how to theorize about truth opened up. On the one hand, there was disagreement about whether theorizing should take the form of an investigation into the formal, logical properties of the truth-predicate, or that of an investigation into the linguistic practices in which the use of such a predicate has a home (i.e. 'fact-stating discourse'). On the other hand, philosophers disagreed as to whether they should be investigating our actual truth-predicate or devising some preferable replacement predicate. Nevertheless, throughout this period, it was almost always a theorist's conception of semantics that dictated the framework in which they investigated truth.

2.2 Logical empiricism (I): Ayer, Carnap, Neurath, Schlick

Like Ramsey, the logical empiricists were deeply influenced by Wittgenstein's *Tractatus* and at least in the first years of the Vienna Circle, and despite their various differences, they adopted a number of that book's fundamental theses, including that of extensionality: that the only propositions with sense are either atomic propositions or those constructed truth-conditionally from them. But they interpreted Wittgenstein's sparse, formal conception of language in the light of pragmatist considerations. For example, they usually read him, tendentiously, as claiming that atomic propositions were in some sense about immediate experience. This conception of the relation of language to experience, together with the thesis of extensionality, led the empiricists to their famous slogan that the only meaningful statements are those that are verifiable. They also inherited one of the major tensions of the *Tractatus*. Both Wittgenstein's picture theory and the verificationist criterion of meaningfulness seem to entail that logical or analytic truths (or falsehoods) are meaningless and thus neither true nor false. On the other hand, these propositions are as truth-functional as any proposition and thus seem apt to be considered true or false.[88] This tension made it difficult for them to formulate a unified conception of truth.

---

[88] For a pellucid description of the tension as it appears in the *Tractatus* (and Wittgenstein's attempt to resolve the tension) see [Fogelin, 1987, 45ff].

The logical empiricists' conception of meaningfulness as verifiability led them to reject a number of traditional philosophical questions as, at least in some sense, meaningless. In particular, because they took questions about the nature of extra-linguistic reality to lead to pointless, irresolvable disputes, semantic questions were impugned too, on the grounds that they concerned mysterious relations between language and this extra-linguistic reality. One of their early projects was thus to create a language suitable for science that would prevent such questions being formulated. To this end, the fundamental expressions of such languages were to be words for the description of our immediate experience. The only other allowable expressions were logical and syntactical expressions and those that could be defined out of the observational expressions using the logical resources of the language. To the extent that semantic questions can be asked in these languages, they must be put in syntactic terms. For example, Carnap suggested that instead of the question 'Is this book about Africa?' we should ask 'Does this book contain the word "Africa?"' [*sic*].[89] In short, for the early logical empiricists the logic of scientific languages was to be the object of study for philosophers and the method was to be purely syntactic.

In the light of this brief characterization, it is not surprising that the popular view of the logical empiricists' theory of truth is that they either rejected truth as a metaphysical pseudo-concept or endorsed a version of the coherence theory of truth by identifying truth with verification or confirmation. The popular view captures some of the positivistic motivations for their various conceptions of truth. Truth was obviously an uncomfortably semantic notion for many of them and such a concept would have to prove both its empiricist credentials and its usefulness in describing relations amongst sentences of a scientific language. The most obvious truth-related concept that met these criteria is verification. But in fact the popular view overlooks the variety of accounts endorsed by the logical positivists. For example, Moritz Schlick shared neither of these attitudes to truth.

Schlick saw himself as holding firm to the Tractarian correspondence theory despite its involving the sort of metaphysical talk that the logical empiricists were trying to purge from serious theorizing – in this case an appeal to both facts and a semantic relation between language and the world. Schlick argued that there was nothing mysterious about the claim that we compare propositions and facts to see whether a proposition is true. Doing so is just what it is to verify a proposition:

> I have often compared propositions to facts; so I had no reason to say that it couldn't be done. I found, for instance, in my Baedeker the statement: "This cathedral has two spires," I was able to compare it with 'reality' by looking at the cathedral, and this comparison convinced me that Baedeker's assertion was true … I meant nothing but a process of this kind when I spoke of testing propositions by comparing them with facts.
>
> [Schlick, 1935, 400]

---

[89] [Carnap, 1935, 65].

Unlike Wittgenstein's theory of truth, Schlick's did not involve the claim that propositions picture reality. Instead, like James and Dewey, Schlick explained the correspondence relation in terms of verification. However, for Schlick verification was understood according to a foundationalist epistemology whose basic propositions were verified by direct confrontation with sense experience. These are captured in protocol sentences 'which express the result of a pure immediate experience without any theoretical addition',[90] so that the experiencer could not make a significant contemporaneous error in judgement. This entails that, once justified, a sentence cannot later become unjustified. According to Schlick, this view also entailed that verification involved directly confronting propositions with facts. This treatment of verification was what the logical empiricists saw as justifying the claim that Schlick's position was a correspondence theory, on the grounds that, on this view, being verified is a *stable* property that propositions have in virtue of comparison with facts.[91] However, given its appeal to extra-linguistic facts and the difficulty it faces in assigning its claims to either the class of empirical or that of syntactic sentences, Schlick's theory of truth was not popular amongst the other members of the Vienna Circle and, with his death in 1936, was entirely abandoned.

At first, Carnap and Neurath shared Schlick's foundationalism. However, under the influence of Duhem and Poincaré, they came to accept confirmation holism and thus abandoned the idea that any proposition could be tested or verified on its own. Instead, they insisted that statements could only meaningfully be compared with other statements. On this view, even basic propositions may well be given up in the face of an inconsistency if this is the simplest way to re-establish a consistent system of belief. They thus replaced Schlick's foundationalist epistemology with a coherence theory of justification. As Hempel emphasized, this move to a holistic view of justification also led them to abandon the thesis of extensionality, for no proposition could be regarded as intrinsically atomic. It also meant that they could not accept Schlick's correspondence theory of truth, for it implies that sentences can change their epistemic status over time and thus contrasts with what they often called the "absolutist" nature of both correspondence truth and our "ordinary" conception of truth. However, *pace* Hempel, their coherence theory of justification does not imply that they were thereby committed to a coherence theory of truth.[92]

That said, this divergence between the "ordinary" notion of truth and their conception of confirmation left Carnap and Neurath with a choice: either abandon the notion of truth altogether, or replace our "ordinary" notion with a notion defined in terms of confirmation. Neurath chose the latter course and so made the move from a coherence theory of justification to a coherence theory of truth:

---

[90] [Hempel, 1935, 11]. See, in particular, Schlick 1934 and 1935.

[91] This identification of what is at issue was anticipated by Bradley [1909, 201–7], in criticism of [Russell, 1907; Stout, 1907].

[92] [Hempel, 1935, 14] seems to follow Neurath in seeing this move as virtually automatic. Walker (1989: ch. IX) is admirably clear that the coherence theory of truth is not implied by the coherence theory of justification.

> If a statement is made, it is to be confronted with the totality of existing statements. If it agrees with them, it is joined to them; if it does not agree, it is called 'untrue' and rejected; or the existing complex of statements of science is modified so that the new statement can be incorporated; the latter decision is mostly taken with hesitation. *There can be no other concept of 'truth' for science.*
>
> [Neurath, 1931, 53] his italics)

In this passage, Neurath not only endorses a coherence theory of truth but suggests that there may be other notions of truth available that have no place in science. In fact, it is clear that Neurath rejected the ordinary notion of truth because it was different from a coherence notion of justification. So while Neurath in a sense offers a coherence theory of truth, he openly acknowledges that this is not the "ordinary" conception of truth. Importantly, unlike Joachim's and Blanshard's, his coherence theory was not based on claims about the essential nature of thought. Nor would Neurath have thought kindly of the metaphysical doctrine that reality is ideally coherent. Rather his holistic, coherence theory of justification together with a distrust of metaphysics led him to a position much like that of the pragmatists.

The case of Carnap is rather more difficult and any interpretation of his pre-Tarski views on truth is bound to be controversial. Many commentators attribute to him a coherence theory of truth based on his coherence theory of justification.[93] But while he may have once agreed with Neurath, most of his published writings on truth in the early thirties present a rather different attitude. In *The Logical Syntax of Language*, despite the remarkably rich notion of syntax to which he appealed, Carnap nevertheless assumed that 'truth' could not be defined syntactically.[94] This was not because he thought that any such definition would lead to contradiction via the liar sentences: he explicitly states a way to avoid the liar paradox by distinguishing clearly between object-languages and metalanguages. Rather, he is clear that once this problem is resolved we have to admit that 'truth and falsehood are not proper syntactical properties; whether a sentence is true or false cannot generally be seen by its design, that is to say, by the kinds and serial order of its symbols.'[95] Of course, he thought that the truth of some sentences, the analytic ones, could be discovered merely by observing their syntactical properties (and one of his ongoing projects was to provide an adequate syntactic definition for the analytic truths). Nevertheless, Carnap is obviously right that no syntactic definition of truth will be possible such that the definition will serve to demarcate the truths from the falsehoods.

Moreover, like Popper,[96] Carnap thought that there was no need for a truth-predicate in logic. He claimed that we could translate the majority of sentences containing 'true' into sentences that do not. In particular, '"*p*" is true' can be translated by '*p*' and 'If "*p*" is true, then so is "*q*"' by '"*q*" is a consequence of "*p*"'. So, for Carnap at this time the truth-predicate was to be avoided primarily because it could not be given an adequate syntactic definition and seemed to lack any legitimate, non-redundant uses. The idea that the truth-predicate is redundant bears an obvious resemblance to Ramsey's redundancy theory of truth. And although neither Carnap nor Popper refers to Ramsey in this context, it is clear that Carnap had read Ramsey and that Popper had read Carnap.[97] However, Carnap did not make the claim that our ordinary truth-predicate was redundant, only that he saw no use for it in logic.

---

[93] See, for example, [Ayer, 1959, 20; Hempel, 1935, 14].

[94] For helpful discussions of Carnap's syntactical resources and his claim that truth could not be defined syntactically, see [Ricketts, 1996; Coffa, 1991, 285–306].

[95] [Carnap, 1934, 216].

[96] [Popper, 1934, 274f].

A. J. Ayer, on the other hand, quite explicitly followed Ramsey in claiming that, owing to the redundancy of the truth-predicate, there was no problem of truth.[98] Ayer, unlike Ramsey, did not turn his mind to the problematic cases for this theory and so offered no further hint as to how they might be handled. However, he did go further than Ramsey in concluding that the correct analysis of truth reveals that there is no property of truth to wonder about the nature of:

> There are sentences ... in which the word "truth" seems to stand for something real; and this leads the speculative philosopher to enquire what this "something" is. Naturally he fails to obtain a satisfactory answer, since his question is illegitimate. For our analysis has shown that the word "truth" does not stand for anything, in the way which such a question requires.
>
> [Ayer, 1936, 89]

It is surprising that a theory of truth so amenable to the logical empiricists was not more generally adopted. In fact, it took the ground breaking logical work of Alfred Tarski to open the eyes of many of the logical positivists to the availability of a theory of truth that was as deflationary as Ramsey's.[99]

2.3 Tarski's semantic conception of truth

Tarski was well aware that the concept of truth was viewed with suspicion by many of his contemporaries. He himself was suspicious of the everyday use of truth-predicates in natural languages. His main concern was that such use led to inconsistency because of the possibility of formulating liar sentences like 'This sentence is false'.[100] One of Tarski's main goals in supplying a definition for a truth-predicate was to ensure that any such term would avoid these paradoxical consequences. Thus, based on his diagnosis of the conditions that allow the liar paradox to be generated for the 'ordinary' truth-predicate, Tarski proposed a number of formal constraints on any adequate definition of a truth-predicate. In particular, he claimed that all semantically closed languages are inconsistent, where a semantically closed language is one that has the means of referring to its own expressions and itself contains semantic expressions like 'true' and 'false'. As well as being semantically closed, Tarski thought that natural languages do not allow us to set up precise formal conditions for the adequate definition of a truth-predicate because we cannot exactly specify the formal structures of these languages (including which of the expressions are meaningful). These two features of natural languages led him to abandon any attempt to define either our ordinary truth-predicate or any other truth-predicate for a natural language. Instead, his project was to define a truth-predicate (or provide a recipe for defining truth-predicates) for formal languages. Moreover, Tarski was only interested in defining the term 'true sentence' for these formal languages and not in any notion of truth as it applies to propositions or beliefs.

---

[97] Carnap refers to Ramsey [1931] a number of times in his [1934] and Popper refers to Carnap's [1934] in his discussion of truth [1934, 275].

[98] [Ayer, 1936, ch. 5].

[99] We do not mean to suggest that Tarski's definition obviously amounts to a deflationary theory.

[100] Amongst the theorists we are considering, Tarski was largely alone at the time in allowing his concern with the semantic paradoxes to shape his views about truth. In these days of paraconsistent logic, it may no longer be as obvious as he assumed that such inconsistencies are an overwhelming cause for concern.

To make sure that the predicate he defined was worthy of being called a *truth*-predicate, Tarski suggested his famous material adequacy condition, Convention T. Convention T asserts that any adequate definition of a truth-predicate, 'T', must entail all instances of the so-called T-schema (T):

$$(T) \quad X \text{ is } T \text{ iff } p$$

where 'X' is replaced by a name or structural description of some sentence and 'p' is replaced by that sentence or a translation of it in the meta-language.[101] (An instance of the schema would thus be "'*Der Schnee ist weiß*' is T iff snow is white", with German as the object- and English as the meta-language.) This is meant to be an obviously legitimate constraint on adequate definitions of truth, as it is meant to be obvious that the related schema (TS) demarcates the set of true sentences.

$$(TS) \quad X \text{ is a true sentence iff } p.$$

So, if a definition of 'T' entails all instances of (T), then 'T' must be at least materially equivalent to 'true sentence'. (TS) is similar to the equivalence emphasized by Ramsey, Carnap and Ayer. It says, in effect, that attributions of truth to sentences are equivalent to assertions of the sentence itself. As well as treating sentences as truth-bearers, however, (TS) also differs in strength from these other proposed equivalences as it is meant only as a material equivalence.

It is important to note, though, that in first offering this condition Tarski rather cautiously claimed that this was only a condition of adequacy on a particular conception of truth, namely the semantical conception, according to which 'a true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so.'[102] Tarski claims that this type of statement is essentially an elaboration on what he calls the '*classical* conception of truth ('true — corresponding with reality').'[103] However, despite his suggestions it is not clear that a semantic definition of truth (and his definition in particular) must amount to a correspondence theory of truth (though it does seem to capture the 'classical', Aristotelian conception).

---

[101] [Tarski, 1933, 187–8].

Finally, it is crucial to Tarski's enterprise that he not use any undefined semantic terms in his definition. Otherwise, the role of primitive concepts would be 'played by concepts which have led to various misunderstandings in the past.' Moreover, 'it would then be difficult to bring this method into harmony with the postulates of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts).'[104] The influence and the vocabulary of the logical positivists are quite plain here: Tarski's goal was not just to avoid the antinomies, but also to provide a definition of truth that would be acceptable to those with positivistic or empiricist suspicions about semantics. Thus he hoped to define semantical concepts 'in terms of the usual concepts of the metalanguage' so that they are 'reduced to purely logical concepts, the concepts of the language being investigated and the specific concepts of the morphology of language.'[105]

Tarski's aim, and his self-imposed constraints, were accordingly these: to produce a definition of 'true sentence' for a formal object-language free from semantic terms and with a precisely specified structure, such that this definition entailed all instances of the T-schema and was formulated in a metalanguage that was about this object-language and used no semantical terms as primitives.

The most obvious way to construct a definition that meets these constraints would be to list all the instances of the T-schema. After all, Tarski himself claims that these instances are partial definitions of truth. For any language with an infinite number of sentences, however, this approach will not provide us with an explicit definition of truth. For such languages the obvious step would be to generalize (TS) to get something like $(T*)$:

$$(T*) \quad \forall p \; (\text{`}p\text{' is a true sentence iff } p).$$

But, as Tarski pointed out, $(T*)$ limits us to naming sentences via quotation names and a more general approach that gives us an explicit definition is (TG).

$$(TG) \quad \forall x \; (x \text{ is a true sentence iff } \exists p \; (x = p \text{ and } p)).$$

As should be clear from our discussion of Ramsey, however, whether we can do this depends on how we read quotation marks and the universal quantifier. If we read the quantifier objectually, then, even if the variables were to range over propositions, the sentence would be ungrammatical. On the other hand, we could read it substitutionally. Substitutional variables do not range over a set of objects so that when we instantiate formulas containing them they are replaced with names of these objects. Instead, they are associated with a class of expressions (in this case sentences) so that when we instantiate we replace the variable with an expression and not a name for the expression. However, whether this approach works will depend on how we read the quotation marks. As Tarski points out, the approach will not work if we treat ‘“p”’ like a syntactically simple name so that the expression inside the quotation marks is no more in quantifiable position than ‘lip’ is in ‘Philip was sleeping’. Rather, we must treat quotation marks as functions that take an expression as input and create a name for it. This allows us to read (TG) so that it not only comes out as grammatical but also seems to provide a definition of truth that meets Tarski's constraints.[106] The only significant reason Tarski gave for not adopting this definition is that he thought that it generates contradiction.[107] Yet if he had been as careful in distinguishing between object-language and metalanguage in this case as he was for his own definition this inconsistency would not arise.[108] There are other reasons for worrying about using substitutional quantification in this context, but Tarski did not share them and we shall see that it is anyway far from clear that they are conclusive.[109]

---

[102] [Tarski, 1933, 155].

[103] ibid. 153. Tarski claims that the semantical definition of truth is an elaboration of this classical conception (and quotes Aristotle's famous slogan about truth as an example of the semantical conception) at 155f.

[104] [Tarski, 1936, 406].

[105] ibid.

Yet another option would be to provide a recursive definition of truth by first defining truth for atomic sentences using the appropriate instances of the T-schema and then defining the circumstances under which complex sentences built out of atomic sentences and logical operators are true. For example, one part of such a definition would be $\langle (A\&B)\rangle$ is true iff $\langle A\rangle$ is true and $\langle B\rangle$ is true.[110] This is obviously much like the logical atomist approach of Russell and Wittgenstein except that the T-schema is used instead of providing a general, correspondence account of truth for atomic sentences. However, Tarski rejected the idea that quantified sentences should be treated like infinite disjunctions or conjunctions and so abandoned the idea that one could define truth for all sentences on the basis of a definition of truth for atomic sentences.

To overcome these problems Tarski began with a recursive definition of a near substitute of sentence truth, namely satisfaction of sentential functions, and showed how we could define sentence truth on the basis of this notion. A sentential function is an open sentence, or a formula that contains variables that are not bound by any quantifiers (e.g. 'x is married to Queen Elizabeth'). To give an idea of what satisfaction is, we can speak of a sentential function's being satisfied by certain objects when replacing the variables with names for these objects results in a true sentence (e.g. Prince Philip satisfies 'x is married to Queen Elizabeth'). However, given that we want to use the notion of satisfaction to define truth, we need some other way to *define* satisfaction. Moreover, Tarski's initial truth definition concerned languages that do not have names, so we will follow him in this and then point out the obvious way to extend his definition. For a finite number of predicates and relations, a simple way to define satisfaction would be merely to list each predicate and relation and the conditions under which it is satisfied (e.g. for all x, x satisfies the predicate 'is married to Queen Elizabeth' iff x is married to Queen Elizabeth).

---

[106] Field offers a similar (but importantly different) approach using schematic variables [Field, 2001, 115 and 141ff].

[107] [Tarski, 1933, 161]. Both Mackie and Soames, for example, agree that this is the only substantial reason Tarski offers. [Mackie, 1973, 32; Soames, 1999, 89].

[108] [Marcus, 1972, 246f]. In Marcus's own words “considerations of definitional adequacy require that for the substitution class of sentences the condition on quantification be
(Q') If $A = (x)B(x)$ then $v(A) = T$ iff $v(B(t)) = T$ for each sentence t such that $(B(t))$ contains fewer quantifiers than A.”

[109] These objections include complaints of circularity (see for example [Platts, 1980, 14f; Horwich, 1998a, 25f], but see also [Camp, 1975; Soames, 1999, 90ff] for responses) and also complaints that substitutional quantifiers are inconsistent with convention T (see, for example, [Wallace, 1971; Kripke, 1976; Camp, 1975] for responses).

[110] We have used angle brackets as a convenient substitute for corner quotes in this sentence.

The task of providing a general definition of satisfaction, however, is harder. Because open sentences can have more than one variable and because they can be closed by applying quantifiers to bind the variables as well, Tarski adopted the technique of talking about sentential functions' being satisfied by *sequences* of objects. A sequence of objects is basically a potentially infinite list of objects where the way the objects are ordered in the list is crucial. When considering whether a sentential function is satisfied by a sequence Tarski pairs the variables in the sentential function with objects in the sequence so that if we have variables x_1, x_2, x_3, they are paired with the first, second and third objects of a sequence. So, if our sentential function were 'x_1 is the husband of x_2' then it would be satisfied by any sequence of objects that had Prince Philip as its first member and Queen Elizabeth as its second (as well as many other sequences, of course). Dealing with the quantifiers is less straightforward, but the general idea is the same. We consider sequences of objects such that the variables of the sentential function are paired with objects in the sequence based on their position in the list.[111]

The general pattern (although not the logical machinery) of Tarski's truth definition is quite easy to grasp. First, he provides a recursive definition of satisfaction of sentential functions by sequences of objects. This crucially involves assigning free variables to objects in the sequences. For sentences there are no free variables and so, whatever sequence we are talking about, this sequence will either satisfy the sentence or not. Once we notice this, the move from a definition of satisfaction to a definition of truth is obvious. True sentences are those that are satisfied by all sequences. False sentences are satisfied by none.

With this ingenious approach, Tarski managed to provide a definition of 'true sentence' that satisfied all the constraints he had imposed. Of course, the definition he supplied is a definition of truth relative to a particular formal language, so his approach could never count as a definition of truth *simpliciter*. That is, as Max Black emphasized,[112] although his work can be applied to different formal languages, the definition of 'true sentence' is unique to the particular language in question. If, as it seems, our ordinary notion of truth is not relative to a language in this way, Tarski has failed to capture the meaning of 'true'. In particular, if we change the language for which we have provided a Tarskian truth-definition by as much as adding one predicate, the definition becomes obsolete. The definition does not tell us how to incorporate this new vocabulary. Moreover, his definitions apply only to a certain sort of *formal* language. Whether or not a Tarski-style definition can be extended to natural languages is of great philosophical significance. One way in which it can be easily extended is to incorporate languages with names. For a language with a finite list of names, a definition of denotation can be given merely by listing the names and the objects they denote. We can then define truth for atomic sentences quite simply along the following lines:

---

[111] For the few who are interested and yet have not met Tarski's treatment of the quantifiers: his treatment of the existential quantifier is such that a sequence s_1 satisfies some existentially quantified sentential function $\exists x_k A$ iff there is some other sequence, s_2, that satisfies A and differs from s_1 at most at the k^{th} place.

> 'Fa' is true iff the object denoted by 'a' satisfies the predicate 'F'.

Of course, as was the case with predicates, any language with an infinite number of names will resist being treated in this way. However, there are also many other types of natural language expressions that resist being treated in a Tarskian fashion; such as indexical expressions, reported speech, adverbial constructions and sentences in non-indicative moods.[113]

So much about Tarski's goals and the extent to which he achieved them is relatively uncontroversial. What is far more difficult to achieve consensus on is what, if anything, Tarski's definition contributes to the philosophical debate about truth. In particular, does it serve to rehabilitate the correspondence theory, as he himself sometimes seems to suggest? Karl Popper certainly thought so, arguing that part of Tarski's great contribution was to point out the adequacy conditions on any correspondence theory of truth.[114] Popper quite rightly sees that a correspondence theory needs the capacity to describe truth-bearers, a way of referring to all the facts described in the object language and an account of the correspondence relation that holds between truth-bearers and facts. But Popper also reads these constraints into Tarski's definition of truth. As we saw, Tarski insisted that the meta-language have the resources to describe or name all sentences of the object language and also to provide translations for all these sentences. For Popper's characterization of Tarski to be right, this latter translation constraint must amount to a way of referring to or describing facts. In other words, as Popper says, the right hand side of the T-schema must state the fact to which the sentence referred to on the left hand side corresponds.[115] However, if we treat facts as whatever is stated by statements like 'Snow is white,' surely we reduce the notion of a fact to something that even coherence or pragmatic theories could admit. There is nothing preventing coherence theorists from accepting that 'Snow is white' is true iff snow is white, provided that they read the sentence as an expression of the belief that snow is white and they commit themselves to the claim that snow is white iff the belief that snow is white is part of the ideally coherent set of beliefs. This is not to agree with Black's claim that Tarski's theory is philosophically insignificant because it is consistent with all philosophical theories. It is simply to point out that Tarski's T-schema (which is not his definition of truth) should be accepted by all theories of truth.

---

[112] [Black, 1948].

[113] Although much work has been done to show how these constructions can be treated in a Tarskian way by those committed to Davidsonian semantics. For some examples of this work see [Davidson, 1980; 1984].

[114] [Popper, 1974, 401].

[115] [Popper, 1974, 402].

Of course, this is also not yet to deny that Tarski's theory might be a correspondence theory. At one time Donald Davidson argued that it was a virtue of Tarski's account that it offered a correspondence theory without needing to appeal to fact-like entities.[116] Tarski's *definition* of 'true sentence' is that a true sentence is satisfied by all sequences. Davidson thus understood sequences as Tarski's replacement for facts. On this reading of Tarski, it is the world that makes sentences true, but it is not the case that each sentence corresponds to some part of the world (a particular fact). It is all the possible sequences of objects taken together that determines that a sentence is true (together with the semantic facts about the sentence). In a sense, then, it is the entire universe that either does or does not make a sentence true. For Davidson, this is just as well. According to Davidson, there is a simple proof, first offered by Frege and now generally known as the 'Slingshot', that shows that any attempt to appeal to facts as the referents of sentences collapses into the claim that there is only one Big Fact that makes all true sentences true.[117] A Tarski-style approach to correspondence theories would also have the rather significant virtue of putting an end to annoying questions about negative facts, general facts, mathematical facts and so on.

However, an appeal to sequences does not on its own constitute a correspondence theory any more than does Tarski's appeal to the availability of metalanguage translations. At a minimum, Tarski's theory would also need to explain the relation that holds between sentences and sequences such that the holding of this relation will constitute the correspondence of the sentence with the world. Davidson and others have claimed that Tarski's definition of satisfaction is precisely what fills this role and so completes the correspondence account of truth. On this reading, Tarski has discovered the relation of satisfaction to be the correspondence relation that philosophers have been searching for. Correspondence thus turns out to be a semantic relation between sentences and the world. In calling such a position a correspondence theory, though, Davidson et al. commit themselves to treating all theories of the following form as correspondence theories:

A sentence is true iff the sentence means that p and p.

This would mean that even Ramsey's redundancy theory is automatically a correspondence theory.[118] In fact, it looks as if, in general, treating Tarski's definition as a version of a correspondence account means that Ramsey's account will need to be treated as a correspondence theory too. Like Tarski, Ramsey noticed that the problem of truth would be easy to solve if all occurrences of 'true' were of the form 'It is true that bitumen is black.' Also like Tarski, Ramsey saw that there were grammatical difficulties in extending this story to cover all cases and that the solution lay in uncovering all the different logical forms that judgements could have and then explaining truth recursively. It is difficult to see what parts of Tarski's superior logical apparatus could turn such a redundancy theory into a correspondence theory.[119]

---

[116] [Davidson, 1969].

[117] There is a large literature concerning the cogency of this argument and the consequences of its conclusion. For an excellent overview and an extensive bibliography see [Neale, 1995].

2.4 Logical empiricism (II): the impact of Tarski

It is worth returning to the logical empiricists to see how the definition of truth given by one of their close associates affected their own conceptions of truth. The difference between Neurath's and Carnap's reactions to Tarski's work is telling. Neurath agreed that Tarski had provided a definition that made sense of our familiar notion of truth rather than of the verificationist notion Neurath himself had been offering. Nevertheless, Neurath continued to worry that this notion of truth carried with it absolutist implications, i.e., that there was both a reality independent of the language in which we describe it and that we could have some sort of absolute certainty in the truth of propositions about this reality.[120] Thus Neurath retained his scepticism towards 'true' until he could be shown that the Tarskian notion of truth was both a useful notion and did not carry such implications.

Carnap's reaction could hardly have been more different. When Tarski described his method for defining truth to Carnap he felt that the scales had fallen from his eyes. Afterwards he embraced semantics and went on to write some of the most influential books on semantics of the twentieth century. Surprisingly, it was not the rather ingenious logical machinery Tarski had used in his definition that Carnap felt was so revolutionary. Rather it was use of the seemingly trivial T-schema as a criterion of material adequacy. For in his [1934] Carnap had most of the logical machinery required for Tarski's definition already in place. Thus in many ways, the step from syntax to semantics was not a major one for Carnap. Yet, for a time, Carnap was blind to the possibility of giving a definition of truth in terms of the conditions that each sentence had to meet in order for it to be true.[121] It took

---

[118] Field [1986] argues that Ramsey's theory is a version of the correspondence theory, based on reading Ramsey as offering a pragmatic account of what makes it the case that judgements have the truth-conditions that they do. According to the way of characterizing truth theories currently under discussion in the text, even if Ramsey had said that there is nothing to say about what makes judgements have truth-conditions he would still have been offering a correspondence theory.

[119] There is much more to be said about the philosophical significance of Tarski's definition. Some of this will emerge as we continue.

[120] See, for example, the last sentence of [Neurath, 1937].

[121] The exact reason for the blindspot is hotly debated amongst commentators. Coffa [1991] claims that Carnap's verificationism led him to look for a syntactic criterion for truth. Ricketts

Tarski to show him the way truth could be so defined and to make him enforce the distinction between truth and confirmation.[122]

The difference between the reactions of Carnap and Neurath shows both that Tarski's T-sentences are helpful in showing which issues are about truth and which issues are not, and also how difficult it can be to heed the lesson. In this case, the T-sentences show quite clearly that truth is not the same as confirmation, that we can abandon the idea of absolute knowledge and yet happily endorse the use of the truth-predicate. For in claiming that 'Snow is white' is true iff snow is white we maintain only that whatever uncertainty there is in the claim that snow is white applies also to the attribution of truth to the sentence 'Snow is white'. Similar things can be said about the independence of reality from the language we use to ask it questions.

2.5 Correspondence vs. redundancy: the Austin/Strawson debate

One of the central issues in the extended debate between J. L. Austin and P. F. Strawson turned on the question of whether yet another issue, namely the theory of meaning, is in fact properly considered part of the theory of truth.[123] As we have seen, Tarski's definition of truth is capable of being turned into a correspondence theory of truth. The fact that it can be so developed stems from its basis in what Tarski calls the semantic conception of truth — a sentence S is true iff S says that some state of affairs obtains and that state of affairs obtains. We saw that an attentive correspondence theorist is likely to claim that this meaning relation which is assumed to hold between the sentence and a state of affairs is the correspondence relation they have been attempting to articulate. Thus the crucial question between a redundancy theorist like Ramsey and a correspondence theorist of this variety is whether an analysis of the meaning relation belongs to the theory of truth.

Austin's correspondence theory seems to be a disguised version of the latter, two-stage, approach to the theory of truth. However, it is important to note that Austin shared the logical positivists' taste for attempting to tackle metaphysical questions through an analysis of language:

> We approach ['truth'] cap and categories in hand: we ask ourselves whether Truth is a substance ... or a quality ... or a relation ... . But philosophers should take something more nearly their own size to strain at. What needs discussing rather is the use, or certain uses, of the word 'true'.
>
> [Austin, 1950, 149]

[1996] claims that it was Carnap's single-mindedness in giving a complete definition of truth that led him to overlook a materially adequate definition.

[122] See [Carnap, 1949] for the classic attack on the conflation. Carnap does not mention his former self as one of those guilty of making the conflation.

[123] The debate between Austin and Strawson began in 1950. It marks the emergence of the continuing feud between deflationary and correspondence theorists.

For Austin, however, the theory of truth should proceed by giving an analysis of our *ordinary* uses of 'true'. The motivation for such an approach is the idea that concepts like *truth* are falsely taken as enigmatic by philosophers who have paid insufficient attention to the way the word 'true' is really used. That said, we will largely ignore Austin's linguistic analyses and merely describe the theory of truth that he ends up endorsing.

Austin complained that previous correspondence theories made two basic mistakes. First, they supposed that the correspondence relation was some sort of real relation between truth-bearers and facts such as an isomorphism or congruence relation.[124] Second, and relatedly, previous correspondence theories incorrectly populated the world with "linguistic *Doppelgänger*" which could stand in such real relations to truth-bearers.[125] In contrast, he urged, the correspondence relation is "*absolutely and purely* conventional."[126] This is because what our words mean is a matter of convention. Once we appreciate this, Austin claims, there is no need to posit facts with the same structure as sentences. Instead, we can say that for a truth-bearer to be true is for there to be some convention that determines that the truth-bearer means that p and for it to be a fact that p. To understand Austin's view in more detail we need to follow him in spelling out the nature of his truth-bearers and conventional relations.

Austin is quite clear that the bearers of truth are neither sentences nor propositions nor beliefs but statements. What is less clear is what statements are meant to be. Austin describes a statement as something that is made (and so is an historic event) in words, but also that one sentence can be used to make different statements (you say 'It is mine' and I say 'It is mine').[127] While it is difficult to be sure, it seems as if Austin means that a statement is an utterance-as-understood, or a pair consisting of an utterance and its interpretation.

Central to Austin's account of the conventional nature of correspondence is the claim that there are two importantly different types of linguistic convention. Descriptive conventions relate sentence types (not statements) to a certain type of state of affairs. Demonstrative conventions relate statements to particular, historic states of affairs. To see the difference between these two conventions, consider a stop sign. There are descriptive conventions that tell us that this sign is a *stop* sign; that the type of thing this sign is telling us to do is stop. But there is also a demonstrative convention that tells us where we should do this — namely in the area that the sign has been planted.[128] Austin's claim is that statements have a similar duality. On the one hand they refer to a particular state of affairs and on the other they assert that this particular state of affairs is of a certain type. This gives us an obvious way of defining truth.

---

[124] [Austin, 1950, 154ff].

[125] ibid. 154.

[126] loc. cit. Austin's italics.

[127] ibid. 151.

[128] Austin suggests the example of traffic signs in [1950, 153, note 10].

> A statement is said to be true when the historic state of affairs to which it is correlated by the demonstrative conventions (the one to which it 'refers') is of a type with which the sentence used in making it is correlated by the descriptive conventions.
>
> [Austin, 1950, 152f]

Another way of putting the same point would be to say that a statement is correlated with a set of states of affairs and a particular state of affairs and the statement is true if the latter is a member of the former. This analysis of correspondence truth also makes a simple account of falsity available that avoids the sort of worries about meaningful falsehoods that Russell struggled with. On Austin's account, statements always refer to actual, real states of affairs. False statements simply assert that they are of a type that they are not.

On the surface, Strawson could hardly have had more to disagree with in Austin's theory. He rejected Austin's account of truth-bearers, truthmakers, the correspondence relation and the types of linguistic convention. As so often, though, their differences were sometimes only apparent.

The key to much of Strawson's dislike for Austin's theory lies in Austin's attempt to elucidate truth using the words 'fact' (or 'states of affairs') and 'statement'. According to Strawson, along with 'true' these words are all part of a linguistic practice, or a particular way we have of communicating with each other, which we can call fact-stating discourse. Strawson's worry is that any attempt to use some of these words to elucidate the others will result in nothing but vacuous truisms. The idea seems to be that one grasps the way to use these inter-definable words all at once as one is introduced into the practice of fact-stating. Strawson's claim is thus that trying to elucidate truth by appealing to these cognate concepts is really an attempt to elucidate truth through elucidating this whole practice. But this elucidation cannot get anywhere if we use notions like 'fact' and 'statement' because if we did so we would find that the 'words occurring in the solution incorporate the problem.'[129] Thus Strawson found it no surprise that Ramsey discovered that if we have first got clear about what a statement (or judgement) is there is no further problem about truth. Presumably Strawson thought that getting clear about the nature of making a statement involves understanding that, within the practice of stating facts, statements aim at the truth.

Strawson's famous attack on the correspondence account of facts exemplifies these rather abstract claims. He began by urging that, *pace* Austin, we do not use statements to refer to facts. If anything, statements *state* facts. However, even if Austin were to give up the claim that statements refer to facts, Strawson argued, there are no such constituents of reality. Facts are just the shadows of true statements. For one, we have no way of individuating facts except for using the statements which the correspondence theory claims are made true by the fact: we can refer to the fact that John is driving only by using the statement that John is driving. It is thus no more of a surprise that statements correspond to facts than it is that there is no further problem of truth once we have understood what it is to make a statement. This is the point of Strawson's oft-quoted remark 'Of course statements and facts fit. They were made for each other.'[130] But it is important to see this claim in the right light. On the one hand, it is meant to suggest the truistic nature of any claims about true statements' corresponding to the facts. But it is also meant to help in throwing doubt on the existence of facts. Because we have no other way of individuating facts except by using the statements that they make true, and because they serve no other role than to be 'the tautological accusatives' of true statements,[131] it seems gratuitous to suppose that such things as facts are part of language-independent reality.

---

[129] [Strawson, 1950, 171].

It is easy to miss the seriousness of this attack. Austin, for one, complained that Strawson had been unfair to facts because we can equally well point out that targets and well-aimed shots are made for each other and yet we do not suppose that targets are bogus entities.[132] In making this parallel Austin rightly argues that it is not sufficient to point out that facts are the internal accusatives of true propositions in order to conclude that the former are the mere shadows of the latter. Yet Strawson's criticism is not meant to rely solely on this point. In the case of targets and well-aimed shots, targets can be set up as the things at which to aim because they can be individuated without having to fire a shot. Further, they can be such that any number of shots are considered well-aimed in virtue of hitting that target. Strawson's worry is that neither of these features is shared by facts. Austin has not shown us how to individuate a fact as the thing to which we should try to get a statement to correspond; nor has he given us reason to think that it is not the case that each true proposition has its own fact that makes it true. However, on this last point, notice that Austin's *statements* (unlike Strawson's) are historical occurrences and so, as for well-aimed shots and targets, any number of them could be made true by the same fact. Here the difference in Austin's and Strawson's choice of truth-bearers seems to have resulted in their talking past each other. For when Strawson claims that facts are the shadows of true statements, he does not mean that they are the shadows of historical occurrences.[133]

This difference between Austin and Strawson leads to Strawson's most serious criticism of the Austinian project. For Strawson claims that Austin's approach rests on a basic confusion between, on the one hand, the conditions that must obtain for our attributions of truth to be true and, on the other, what we actually assert when we make an attribution of truth. By focusing on the first issue, Strawson claims that Austin is led to an attempt at elucidating fact-stating discourse by disclosing the conventional relations between our statements and the world. But, according to Strawson, the real issue is what we assert in making an attribution of truth and we do not assert that such conventional relations hold.[134] Instead, Strawson says, we should focus on the role of 'true' within fact-stating discourse and not attempt to step outside this type of discourse to explain truth. In saying this, Strawson repeats an objection to the correspondence theory that we have found being made by Joachim, Blanshard, the pragmatists and the logical positivists – namely, that the correspondence theory attempts to step outside language or the mind to compare language to the world when the right (and only possible) approach is to attempt to understand truth from within this discourse. We will see Quine make the same sort of remark.

---

[130] ibid. 168.

[131] [Armstrong, 1997, 19].

[132] Austin 1961: 188.

[133] Strawson spells out his conception of statement at [1950, 162–5].

Strawson's conclusion, then, is that Austin has illicitly included claims about the nature of representation into the discussion of truth. As such, Strawson's attack clearly generalizes to all theories that attempt to explain correspondence by focusing on representation. Notice, though, that the difference between Austin and Strawson on the role of the theory of meaning may well reduce to the differences in their choice of truth-bearers. For if Strawson is right and we should treat propositions as the bearers of truth, then it seems to be entirely superfluous to add a discussion of how it is that our sentences get to express the propositions that they do. However, if Austin was focusing on sentences as the bearers of truth then perhaps a case could be made for the inclusion of such a discussion, as Strawson appeared to recognize:

> If someone wishes to contend that we do not *really*, or do not *fully*, know the meaning of 'is true' unless we know what types of conventional relation obtain between words and things when something true is stated or otherwise expressed in words, then the contention seems to me by no means extravagant. ...
>
> Better, perhaps, let the theory of truth become, as it has shown so pronounced a historical tendency to become, part of some other theory: that of knowledge; or of mind; or of meaning.
>
> [Strawson, 1964, 232 and 233]

Here Strawson articulates another common theme, this time that of the deflationist. Those who follow Ramsey's approach to truth see the theory of truth as fairly straightforward. It is only by mixing up truth with other topics that we come to see it as problematic.

In fact, Strawson himself championed a Ramsey-style approach to the nature of truth. Focusing on how we use 'true', Strawson agreed with Ramsey that to state that p is true is to make no further statement than would be made by stating that p, but unlike his predecessor Strawson made much of the idea that we use 'p' and '"p" is true' differently.[135] He claimed that sentences containing

---

[134] ibid. 172ff.

[135] Ramsey [1923, 142] said only that we use 'true' for different emotive or stylistic reasons or to indicate the position of the proposition in an argument. Carnap [1942, 26] also claims that

'true' are much like performatives such as 'I promise to clean the house.' So that although we assert nothing more than that p with the statement that p is true, we do more than just assert that p. We also perform the speech act of endorsing or confirming someone's statement. According to Strawson, attributions of truth require someone to have first made the statement to which we are attributing truth so that we can confirm or endorse it. In his earlier presentations of this view [1949; 1950] Strawson argued that this fact tempts people to the view that we are attributing a property to a statement. However, we are not attributing truth to anything, we merely require an appropriate context, namely someone's making an assertion, before we can signal our agreement. This view can look as if it may be required by any Ramsey-style, or deflationary, theory of truth. For if "'p' is true" asserts no more than 'p', then, given that 'p' is not about a statement, our "attributions" of truth should not be either.

Austin, for one, rejected the idea that we do not use 'true' to say something about statements. The rejection is *prima facie* reasonable, as the move from (1) and (2) to (3) seems natural:

1. John said that Brian is sleeping.
2. It is true that Brian is sleeping.
3. Something John said is true.

But how could we make sense of this style of reasoning if (2) was not about anything at all? In any event, Strawson backed away from this view in later writings. In his [1964], for example, he went to some length to argue that deflationary theories are compatible with the claim that attributions of truth are in fact statements about statements. He suggested that we make the following paraphrases of statements that attribute truth:

4a. John's statement that p is true.

4b. As John stated, p.

5a. It is true that p.

5b. As may be urged or objected or ..., p.

Strawson's hope in offering these paraphrases was that he could provide us with statements that (i) did not use 'true', (ii) were equivalent to statements that did use 'true' and (iii) were about (in some sense of 'about') statements. (5b) in particular, however, seems a strange paraphrase of (5a) precisely because (5a) does not seem to be making a claim about what may be urged or objected or... Yet for Strawson's performative approach to work, (5a) must be read in such a

---

while 'p' and "'p' is true" mean the same in a logical or semantic sense, the two expressions have 'different features and different conditions of application; from this point of view we may e.g. point to the difference between these two statements in emphasis and emotional function.'

way because Strawson insists that such attributions, if they are about something, are about actual or possible acts of stating.

Strawson's comments about the role of 'true' can be seen as a suggestion about the utility of the truth-predicate. If attributions of truth say no more than the statement to which they attribute truth, then why do we have a truth-predicate at all? Strawson's suggestion is that the predicate is useful as a performative device. One question that should be asked of any such approach is whether it can do better than Ramsey's in dealing with the use of 'true' in statements like 'Everything the Pope says is true'. If 'true' is regarded purely as a performative device, it is difficult to see how to read such statements. Strawson's original suggestion was something like (6):

6. The Pope has made some statements. I confirm them all.

But Strawson's view is that we do not actually *assert* that we confirm them; rather we just confirm them. This entails that the only thing asserted by (6) is that the Pope has made some statements, which of course is true. But, as Soames points out, this means that even if some things the Pope says are false, according to Strawson's analysis the statement 'Everything the Pope says is true' will come out true.[136] In the light of such difficulties, Strawson later altered his theory by de-emphasizing the confirming use of 'true' as only one use among others. The analysis he later gives of these problematic statements seems to follow Quine.

2.6 Quine and disquotation

Like those of the logical empiricists, Quine's views on truth were shaped by a distrust of semantic concepts. In particular, Quine is famous for his strong claim that there is no such thing as objective, interlinguistic synonymy, and that it only makes sense to compare the meanings of expressions that belong to the same language. He was led to this conclusion by his denial of meaning atomism, which, like Joachim and Blanshard, Quine used to reject the notion of individual propositions associated with individual sentences. However, if there is no isolable meaning for each sentence, then the correspondence theory is threatened: we cannot say that a sentence is true iff it represents some particular state of affairs as obtaining and that state of affairs obtains.

Continuing his agreement with the early coherence theorists, Quine insisted that only whole theories can be compared with reality.[137] This confirmation holism was a crucial feature of his attack on the distinction between analytic and synthetic truths.[138] No truth, he claimed, was immune to revision in the event of a theory's being confronted with a recalcitrant experience. The difference between the sort of statements that had been labelled analytic and those that had been labelled synthetic was thus merely one of degree. Although this attack on the analytic/synthetic distinction was largely directed at the logical empiricists, removing this distinction from their theories of truth would have removed a worrying tension in their views that we remarked on earlier (§2.2, opening paragraph), namely, that between their truth-functional account of molecular propositions and their claim that analytic propositions are neither true nor false.

---

[136] [Soames, 1999, 236f].

[137] [Quine, 1970, 1–8], spells this out clearly.

[138] See [Quine, 1953].

Yet Quine's denial of meaning atomism also led him to reject epistemic theories of truth like coherence and pragmatic theories. As we have seen, such theories are constantly in danger of succumbing to relativism. To avoid this, their proponents often resort to the idea of an ideal system of beliefs (perhaps the set of beliefs we would adopt at the end of inquiry or perhaps that ideally coherent set that is identical with reality). Any particular belief is then said to be true iff it is a member of this ideal set. As Quine points out, a serious problem arises if we combine this approach with meaning holism (as Joachim and Blanshard did). For if we cannot ask about synonymy interlinguistically (intertheoretically), then we cannot make sense of the idea that a belief that belongs to our current theory also belongs to some ideal theory.[139] For no two beliefs (sentences) that belong to different sets of beliefs (theories) can have the same content. Thus meaning holism undermines epistemic theories of truth.[140]

In contrast to both coherence and correspondence theorists, Quine urged that truth was 'immanent'. In other words, the truth-predicate can be meaningfully applied only to sentences within the speaker's own language. He argued that to the extent that we can call foreign sentences true, this is only relative to some translation scheme that we employ to translate these sentences into sentences of our language. Yet, once truth is treated as internal to a language and in consequence (because Quine thought that people with different theories thereby speak different languages), as internal to a theory, problems about its nature dissolve in the way a deflationary reading of Tarski would suggest. With an empiricist's distrust of propositions, Quine takes sentences as truth-bearers, though, to avoid problems concerning indexicals and other context-dependent expressions, he is careful to define a class of sentences he calls 'eternal', which function as the primary bearers of truth. Eternal sentences are sentence types whose tokens always have the same truth-value. For example, while 'I am hungry' is not an eternal sentence, 'James Brown is hungry as at 10.15 p.m. on 13^{th} August 1968' supposedly is. For these sentences, Quine claimed that truth is disquotational. An attribution of truth to a sentence merely undoes the effects of quotation marks that we have used to form a name for the sentence. As Ramsey noted, this means that in simple contexts the truth-predicate is redundant. This is seen in Tarski's T-schema, and more clearly in what has come to be called the disquotation schema:

$$(DS) \quad \text{`}P\text{' is true iff } P.$$

---

[139] [Quine, 1960, 23f].

[140] Of course, a pragmatist or coherentist could accept the charge of relativism and thus accept both meaning holism and an epistemic theory of truth. Rorty sometimes seems to be happy to adopt this position, for example.

Whereas the original T-schema allowed an arbitrary name for the sentence, the disquotation schema requires that the name be formed by the addition of quotation marks.

Where Ramsey seems to have gone wrong is in not noticing that the difficult cases for the redundancy view actually show the *raison d'être* of the truth-predicate. For in cases of generalization (like 'Everything the Pope says is true' or "All instances of 'If p, then p' are true") we cannot dispense with the truth-predicate precisely because these are the cases in which it does its work. In these cases we want to generalize in a way analogous to the generalization from a sentence like 'Socrates is mortal' to get 'For all x, if x is a man, then x is mortal'. However, we cannot do this because the move from 'If time flies then time flies' to 'For all x, if x then x' ends with a string of expressions that is incoherent if the quantifier is read objectually. The beauty of the truth-predicate is that it allows us to make these generalizations:

> We could not generalize as in 'All men are mortal', because 'time flies' is not, like 'Socrates', a name of one of a range of objects (men) over which to generalize. We cleared this obstacle by *semantic ascent*: by ascending to a level where there were indeed objects over which to generalize, namely linguistic objects, sentences.
>
> [Quine, 1990, 81] his italics)

Because 'x' is equivalent to "x' is true', 'If time flies then time flies' is equivalent to 'If 'time flies' is true, then 'time flies' is true'. But the final sentence can be used to create a coherent generalization: in fact, just the generalization which seemed at first to pose a problem for a disquotational view of truth. According to Quine it is thus the transparent or disquotational function of 'true' that allows us to make a technical, semantic ascent to talk about sentences, while still talking about the world. With this suggestion about the utility of the truth-predicate, Quine's disquotationalism purports to solve two of the more difficult problems facing deflationary accounts of truth. On the one hand, it explains why it is we have a truth-predicate if its uses are often redundant. On the other, Quine's explanation of the utility of the predicate explains how 'true' functions in precisely those cases of generalization that proved a stumbling block for both Ramsey and Strawson.

Thus disquotationalists treat 'true' as a meta-linguistic predicate applying to sentences. Doing this allows us to explain generalization contexts without appealing to any nonstandard understanding of the quantifiers. However, as Quine essentially admits by focusing on eternal sentences, disquotationalism has difficulty with a vast range of sentences of natural language. For example, treating "'I am hungry' is true" disquotationally implies that it is equivalent to 'I am hungry'. Of course, this implication means that we cannot use the disquotational account to describe the conditions under which some other person's utterance of 'I am hungry' is true. In attempting to do so, we would be forced to say that their utterance is true iff *I* am hungry. Another difficulty arises if the disquotation schema is treated as expressing a necessary equivalence, as seems required if disquotationalism is to be in conflict with inflationary theories such as correspondence or coherence. This reading of the schema implies that the truth-conditions of sentences belong to them necessarily. In other words, in a possible world where we used the sentence 'Rabbits are furry' radically differently, and so the sentence's meaning differed from that in this world, the sentence would still be true iff rabbits are furry in that world. Intuitively, though, it seems that sentences, as opposed to propositions, have their truth-conditions contingently. These problems, and the closely related problems associated with the disquotationalist's claim that 'true' is confined to a particular language, suggest that the simple disquotationalist account will need to be modified if it is to capture the ordinary meaning of 'true'.[141]

2.7 Dummett

Michael Dummett, however, raised a number of further difficulties for the deflationary approach in the course of mounting a defence of a verificationist theory of truth inspired by intuitionist and constructivist accounts of mathematics.[142]

Dummett expressed his reservations about deflationary theories of truth as concerns about the attempt to define truth in terms of the equivalence thesis:

$$(ET) \quad \langle \text{It is true that } A \text{ iff } A \rangle.$$[143]

Each of these concerns has been influential in the ongoing debate about deflationary theories of truth. But each rests on some controversial, though not necessarily implausible, assumptions. First, Dummett argued that the equivalence thesis runs into contradiction if we deny the law of bivalence.[144] The simple argument is that if a statement, say A, is neither true nor false, then it is false that A is true. However, this means that while $\langle A \rangle$ is neither true nor false, $\langle$it is true that $A\rangle$ is false, and so $\langle A \rangle$ and $\langle$it is true that $A\rangle$ are not equivalent. This argument draws on the obvious inconsistency between the following:

7. Some statements are neither true nor false (Denial of Bivalence)
8. For all statements A, A is true iff A (Equivalence Thesis)
9. For all statements A, A is false iff not A.

---

[141] See [David, 1994, ch. 5], for a discussion of the more pressing problems facing the disquotationalist.

[142] [Dummett, 1959]. Although he ends this paper by rather strangely suggesting that he adhered to a version of the redundancy theory, the commitment fitted badly with much of the rest of that paper. He later admitted that this way of characterizing his position was misleading. See his [1978, xxii].

[143] [Dummett, 1973, 445]. Dummett uses corner quotes to express the thesis. For convenience, we have substituted angle brackets.

[144] [Dummett, 1959, 4–7]. See also [Dummett, 1973, 445–446].

Of course, whether or not someone like Quine should worry about this inconsistency depends on whether one thinks that (9) reflects an adequate conception of falsity and whether one thinks that we should abandon bivalence. We shall explore Dummett's verificationist reasons for abandoning bivalence below.[145]

Dummett's second influential criticism is the now familiar one that the redundancy theory does not account for truth's being a normative goal of assertion.[146] And he insists that the redundancy theory cannot be fixed by simply adding the claim that in making an assertion we aim at the truth. What he requires by way of supplementation 'is a description of the linguistic activity of making assertions; and this is a task of enormous complexity.'[147] In other words, he is demanding what Strawson denied was the task of a theory of truth, namely the analysis of fact-stating discourse. However, it is worth recalling from §1.6 that Ramsey had already offered an explanation of the normative role of truth that is consistent with a redundancy theory of truth. If one provides a version of a 'success semantics' then it is quite clear why we aim at the truth: true beliefs are those that lead to successful action. Dummett's demand had been met before it had been made.

Dummett's concern that the activity of asserting has been neglected leads fairly directly to his third influential claim about the redundancy theory. He argued (or rather claimed it was too obvious to be able to argue) that the redundancy theory was incompatible with truth-conditional theories of meaning; that is, theories of meaning that took truth as their fundamental concept. Traditionally, he claimed, theories of truth have tried to do more than spell out the conditions under which assertions are true. They have tried to give an account of truth that would serve as the basis for a theory of meaning. However, if one tries to explain 'true' by appealing to instances of the equivalence thesis, one cannot then use these instances of that thesis to explain the meaning of the sentences. For one can understand the explanation of 'true' only if one already understands the sentence that is used on the right hand side of the equivalence thesis:

> But in order that someone should gain from the explanation that P is true in such-and-such circumstances an understanding of the sense of P, he must already know what it means to say of P that it is true. If when he enquires into this he is told that the only explanation is that to say that P is true is the same as to assert P, it will follow that in order to understand what is meant by saying that P is true, he must already know the sense of asserting P, which was precisely what was supposed to be being explained to him.
>
> [Dummett, 1959, 7]

Dummett has argued that this problem is especially apparent if we consider a Tarskian truth-definition in which the metalanguage is not an extension of the object-language.[148] If we do not understand the object-language, then this sort of truth-definition will not help us understand its sentences because we will have no idea what is accomplished by pairing them with truth-conditions: the appeal is to his second objection, that deflationary theories cannot tell us the point of calling sentences true. Furthermore, these theories seemingly assign truth-conditions to sentences without basing these assignments on the way each sentence is used. This purported feature of deflationary theories of truth-conditions would make them implausible theories of meaning.

---

[145] Notice that Dummett only thinks we should abandon bivalence and not deny it. He distinguishes between (7) and the rejection of the claim that every statement is either true or false.

[146] [Dummett, 1959, 2f].

[147] [Dummett, 1978, 20].

Dummett's conception of theories of meaning is crucial to his famous argument(s) for anti-realism. However, the connections between this conception and his conception of realism are tortuous; our account must be brutally brief. He characterizes realism as the belief that statements 'possess an objective truth-value, independently of our manner of knowing it: they are true or false in virtue of a reality existing independently of us.'[149] Dummett thus treats realism as a doctrine about truth, namely, that it is recognition-transcendent. He also urges that this recognition-transcendence is necessary to guarantee the principle of bivalence. That is, in cases where we have no means of determining whether a sentence is true or not (where no condition of which we are capable of becoming aware determines whether or not the sentence is true), it is only if there is some recognition-transcendent condition which, if met, determines that the sentence is true (and if not that it is false) that the sentence must be either true or false. In other words, without recognition-transcendence, bivalence, and hence the objectivity of truth, is at risk.[150]

Without a guarantee for the principle of bivalence, however, so Dummett argues, we cannot adopt a truth-conditional semantics. He argues at length that the nature of assertion is such that we only speak of an assertion's being correct or incorrect and have no need and no room for the idea that there is a third possibility that lies between these poles.[151] The reason he concentrates on this supposed feature of assertion is this: if the principle of bivalence cannot be guaranteed, and so when it comes to truth there can be more than two possibilities to consider, then truth-conditional semantics cannot be the right approach to the theory of meaning. (At least, not if truth-conditions are understood recognition-transcendentally.) Thus, Dummett suggests, realist/anti-realist debates about a particular subject matter ought to be reformulated as debates about what notion of meaning, and, because meaning should be understood truth-conditionally, what notion of truth is appropriate for a particular area of discourse.[152] Accordingly, the question as to the nature of truth-in-general is better formulated as 'What is the notion of truth appropriate for this particular area of discourse?' We shall see that the pluralism about the nature of truth implied by this line of thought has been developed more recently by Crispin Wright and others.

---

[148] [Dummett, 1978, xxi].

[149] ibid. 146.

[150] [Dummett, 1959, 14ff].

[151] [Dummett, 1959]. This is not to suppose that Dummett rejects outright the possibility of three truth-values. He claims that we may need to appeal to more than two truth-values to explain the way the meaning of complex sentences depends on the meaning of atomic sentences. However, he later [1973, 446f] distinguishes between the "ingredient sense" (the sense relevant to its contribution to more complex sentences) of a sentence and its content. As far as a sentence's content is concerned only two possibilities can be of concern.

[152] The situation is a bit delicate here. Dummett argues that if we give up recognition-

The general structure of Dummett's argument for anti-realism is that a recognition-transcendent notion of truth-conditions is untenable.[153] We have already seen that he conceives a theory of meaning as a theory that specifies for each sentence the conditions under which the understanding of the sentence counts as being manifested. If the obtaining of truth-conditions of a sentence is beyond our recognitional capacity, he claims, we could never manifest our understanding of these sentences (the manifestation argument). Moreover, we could never be taught the meaning of these sentences to begin with (the acquisition argument).[154] He concludes that we should adopt instead a notion of truth that is relative to our epistemic powers: in short, a verificationist conception of truth. The right notion of truth-conditions to use for a theory of meaning is therefore a verificationist notion of truth-conditions.[155] This allows us to abandon bivalence, because some statements will be neither verified nor have their negations verified. However, an assertion will still only be either correct or incorrect, because every assertion will be either verified or not and we can thus say an assertion is correct if and only if it has been verified. The compatibility of these two claims is ensured by the fact that, unlike truth, verification is such that the fact that a statement is not verified does not entail that its negation is verified. For example, although we have not verified the claim that there is life on the planet Venus, this certainly does not entail that we have verified that there is not life on Venus.

Dummett's arguments for verificationism about truth have been repeatedly criticized. It is worth pointing out just a few of the more salient of their controversial assumptions. For one, his acquisition argument relies on the assumption that to learn the meaning of a sentence we must be acquainted with the conditions under which it is *conclusively* verified.[156] Dummett here adopts an extreme version of verificationism about meaning that even the logical positivists were forced to abandon. He himself relinquished this position too,[157] but it is unclear how his argument survives without it. He also rejects Quine's arguments against meaning atomism. For it is crucial to Dummett's manifestation argument that each statement can be treated as having its own verification conditions and that a theory

---

transcendent truth-conditions we must give up the idea that truth is the *central notion* in a theory of meaning. However, identifying truth with verification still allows us to treat meaning as truth-conditional in the sense that meaning is identical with verification conditions. More on this below.

[153] See [Dummett, 1976] for an extended discussion of his conception of theories of meaning.

[154] A brief version of this line of thought is presented at [1959, 17]. See also [Dummett, 1973, 467f].

[155] Although Dummett sometimes characterized his claim as being that we should abandon truth-conditional theories of meaning, he later realized that what he meant was that we need to shift conceptions of what a truth-condition is because we need to treat truth as verification.

[156] This is made clear at Dummett 1976: 132.

[157] [Dummett, 1978, xxxviii].

of meaning can spell out the conditions under which understanding of a particular sentence can be manifested. Whatever the merits of Quine's arguments in general, it is difficult to see how one can deflect their force once the verificationist background is accepted.

Finally, it is worth noting that verificationists about meaning are not *ipso facto* verificationists about truth. In particular, even if Dummett is right that the meaning of a statement is the conditions under which it is verified, this entails a claim about truth only if we take the meaning of a statement to be the conditions under which it is true. For example, there is nothing preventing a verificationist about meaning from following Ayer in rejecting truth-conditional theories of meaning and embracing a redundancy theory of truth. Nor is it clear that Dummett's arguments for anti-realism should be construed, as he does construe them, as arguments about the nature of truth. Although we cannot argue for it here, it seems to us, and to many others, that realism (or anti-realism) is a metaphysical addition to semantics and not built into it.[158] This seems to have been Quine's point when he claimed that from within the language there is nothing more to say about truth than to point out its disquotational properties. Perhaps, then, the outstanding question is whether the right perspective to take on the nature of truth is from inside or outside our language. Indeed, this question proved important to a number of theorists of truth in the latter stages of the century.

3 LATER VIEWS

3.1 Introductory comments

By the early 1970s the linguistic approach to philosophy was beginning to fall from favour. In consequence, theorists of truth began to lose confidence in the idea that the everyday behaviour of the predicate 'true' is the key to the nature of truth. Yet, as the matters were separated, both were considered legitimate parts of a theory of truth. This tolerance is reflected in the abundance and diversity of theories that blossomed in the last decades of the century. And while many of these accounts are clearly descendants of those already encountered, the evolutionary distance between them is sometimes large. To further compound the difficulty of providing a succinct, coherent narrative for this era, at this time, more than ever, philosophy's characteristically looping method of progression resulted in "earlier" discussions being regarded by later writers as equally contemporary as those far closer to them in actual date.

3.2 Correspondence without facts: Field

Most philosophers who wrote on truth in the later twentieth century saw themselves as either responding to or developing Tarski's formal definition of truth. In his seminal paper of 1972, Hartry Field aimed to do both. In Tarski's work

---

[158] For a brief and clear argument to this effect see [Soames, 1999, 32–9].

Field saw the potential for a powerful correspondence theory liberated from the need to appeal to facts. However, according to Field, Tarski's definition of 'true sentence' failed to satisfy two plausible constraints on a theory of truth: first, it should capture the meaning of 'true', and second, it should give a physicalistically respectable account of the nature of truth.[159] Because of the language-relative nature of Tarski's definitions, the first is not satisfied; and the second is not met because (despite his advertisements to the contrary) Tarski had not managed to provide a definition of truth free of semantical terms. Even though Tarski himself may not have been concerned to be "physicalistically respectable" in Field's sense, both objections are serious in ways which he would have had to recognize, and it is worth looking at Field's arguments for them.

The two defects of Tarski's definition stem from the same source — the list approach to defining satisfaction (and reference). As we saw (§2.3), basing the definition of truth on these lists means that it will not be applicable to new languages, or even the same language with as little as one new name added. Yet, Field argued, even if we ignore this defect, the list approach goes against one of the basic tenets of physicalism.[160] To show this, he relied on an analogy with the explanation of numerous chemical facts by appeal to the property of elements known as 'valency'. Each element has a certain valency which determines how it can combine with other elements to form molecules. As Field points out, one way to explain what valency is would be to list each element along with its corresponding valency. Such an explanation should not satisfy a physicalist looking for an account of the nature of chemical valency, however, for the explanation does not even hint at the physical basis of these chemical properties. It merely postulates these properties as primitives. According to Field, it is only because we learnt that chemical valencies can be reduced to the electron configuration of atoms that these chemical properties have shown themselves to be physicalistically respectable. Similarly, a physicalist should not be content with Tarski's list of names and their referents, and predicates and their extensions. What is wanted is an explanation of what physical facts underlie these semantic facts. Without such an explanation, Tarski has only shown us how to reduce truth to other semantic concepts.

Field accordingly suggests that a satisfactory physicalist theory of truth would involve two stages. First, we use Tarski's work to define truth in terms of reference and satisfaction, of which we then give a physicalistically respectable theory. In other words, Tarski's definition is to be supplemented with a physicalistic story about what makes it the case that primitive terms have the semantic values they do.[161] It is important to recognize, however, that Field should have also demanded that a theory of truth provide an account of what makes it the case that the elements of the logical vocabulary have the semantic values they do. For it should be equally anathema to a physicalist to either assume or stipulate that some syntactic form is a negation as it is to stipulate that some name refers to some person.[162] But though accepting this point may make the position more difficult to defend, it changes little in the general approach.[163]

---

[159] It is characteristic of much late twentieth century analytical philosophy that physicalism is treated as axiomatically true, so that debates tended to concern merely which version is to be preferred.

[160] [Field, 1972, 15–22].

[161] Field 1972: 19, gives a causal theory of reference based on Kripke 1972 as an example of such a theory, but it is clear that any theory that explained the physical basis of reference would be satisfactory. Notice that not just any physicalistically respectable theory of reference will yield a correspondence theory of truth in Field's sense, because a theory of reference that paralleled a

Field claimed that the result would be a correspondence theory. Unsurprisingly this sort of correspondence theory has proved popular.[164] It offers to rehabilitate a traditional and intuitive account of truth shorn of many of its implausible accretions. Yet although Field hoped to avoid some of the defects of Tarski's definition, he inherits one already mentioned: it is extremely difficult to apply it to many types of natural language constructions (see §2.3). Further, Field's fact-free correspondence theory relies on a workable theory of primitive denotation, a theory that three decades of philosophical research has failed to uncover. Moreover, three fundamental concerns remain.

First, one might hesitate in considering such a theory to be genuinely one of *correspondence* since both facts and any fact/utterance correspondence relation have been replaced in it by mere word/thing relations. However, Austin's explanation of the correspondence relation consisted in the elucidation of *meaning*, i.e. what makes '*p*' mean that *p*. Similarly, Field hoped to use Tarski's work to turn an account of reference and satisfaction into an account of what makes it the case that a sentence has certain truth-conditions. And, of course, a sentence is true when its truth-conditions are satisfied.[165]

Second, the comparison with Austin's theory should remind us of Strawson's complaint that it is difficult to see why such a theory about meaning and representation adds anything to the theory of truth, especially if truth-bearers have their truth-conditions essentially. Although Field has taken sentence-tokens as truth-bearers, Soames has argued that sentence-tokens are unable to play this role in a Tarskian truth definition, pointing out that there is no general principle for determining, for example, when one sentence-token is a negation of another. Instead, he urges, we should treat sentence types, abstract entities that belong to abstract languages and have their truth-conditions essentially, as truth-bearers.[166]

A third problem for Field concerns his valency analogy. In general, a physicalist will only require the sort of reduction available in the case of chemical valency when the property in question plays a causal-explanatory role in our best theories. A

---

deflationary theory of truth would be acceptable to a physicalist.

[162] [Soames, 1984, 405–8; Stalnaker, 1984, 30f].

[163] Field has since agreed with Soames and Stalnaker, but downplayed the importance of their observation. He points out that the general approach merely needs to be supplemented with a theory of reference for the logical vocabulary. This supplementation can seem impossible to come by if one assumes that such a theory must be a causal theory, but there is no reason to assume this. [Field, 2001, 27].

[164] For a defence of this approach see [Devitt, 1991].

[165] Field elaborates on this approach to truth in the opening paragraphs of Field 1986.

[166] [Soames, 1984, 410ff].

property that does not help to explain anything, or does non-causal explanatory work, should not be expected to be susceptible to such a reduction. Thus, if Field's analogy is to work, then truth must be a causal-explanatory property.[167] This realization led to a search for contexts in which truth plays a causal-explanatory role in the hope of thereby finding a motivation for the correspondence theory.

One realm in which truth might be taken to be explanatory is the theory of meaning. Davidson, for example, saw Tarski's truth-definition, because of its recursive character, as the best way to give a theory of meaning for a language.[168] Such a truth-definition tells us that certain utterances are true under certain circumstances. It was Davidson's contention that this information gives the meaning of those utterances, so that knowledge of the theory would give a finite being understanding of the language despite the potential infinity of utterances within it.

However, recall that Dummett insisted that Tarski's original definition cannot also be used to explain the meanings of the expressions of the language. The definition relies on our knowing these already. Of course, we can treat Tarski's definition in the opposite way and use it to explain the meaning of the terms of the language on the assumption that we already grasp the notion of truth that appears in the theory. Once Davidson realized that one had to choose between these two projects, he declared that he meant to be using the concept of truth to explain meaning and thereby the behaviour of those who use the language for which we have provided a theory of meaning.[169] It follows that, if Davidson is right about the form of theories of meaning, then truth has an explanatory role: it is the (or a) central concept in the theory that we use to explain behaviour. Unfortunately, Davidson's approach is incompatible with Field's for the reasons just canvassed. Field wanted to explain truth through providing a theory of content. Davidson used truth in order to provide a theory of content. Nevertheless, what Davidson's theory highlights is that we use content attributions in order to understand and to explain behaviour. *If* these attributions are essentially attributions of truth-conditions, then it looks as if truth is a central notion in an explanatory theory. Indeed, in Field's later work on truth, he has regarded the fundamental question as being whether what is causal-explanatory is, not truth itself, but rather, the property of possessing a given set of truth-conditions.

A different context in which truth seems to play an explanatory role is that of explanation of success. Whether it be the success of science or of individual agents in attaining what they want, we often explain success in terms of the truth of certain hypotheses or beliefs.[170] If someone's attempt is successful because he had a true belief about how to achieve his goal, then why not say that his success is due to the truth of his belief? In his original article Field said similar things about the role of truth in learning from others.[171] It seems plausible that, for many people we interact with, the explanation of why we believe what they say is that they generally speak the truth. So, even if we do not follow Field in seeing the debate over the explanatory role of attributions of truth-conditions as relevant, there is also a *prima facie* case for supposing that *truth* plays a causal-explanatory role. Perhaps, then, the strictures of physicalism will require a reduction of truth to physical properties parallelling the reduction of chemical valencies to electron configurations.

---

[167] This point has been pressed by Leeds and, following him, Putnam [Leeds, 1978; Putnam, 1978, Lectures I and II].

[168] [Davidson, 1967].

[169] [Davidson, 1990, 286].

[170] [Putnam, 1978; Field 1986]. The view is obviously inspired by Ramsey [1927], but overlooks the fact that Ramsey's view is compatible with a redundancy theory: §1.6 above.

3.3 Redundancy without redundancy: Grover, Leeds, Prior, Williams

A year before Field argued for the substantiality of the property of truth, Arthur Prior was arguing for an extreme version of the opposite view. Like Ayer, he maintained that truth was not a property at all, but added that '... is true' should not be treated like a predicate and that sentences involving 'true' are equivalent to sentences that are 'true'-free. In this sense, Prior's treatment of truth, and Dorothy Grover's development of it, are redundancy theories. Yet both Prior and Grover also maintain that constructions involving 'true' perform useful functions. In this sense they do not suppose 'true' to be redundant at all. Whatever the label, their approach has a number of virtues, not least of which is that it suggests a neat general response to arguments, of which Field's is an example, that truth is a substantial property.

Prior begins with the idea that the primary truth-bearers are whatever the objects of thought are. As he maintains that it would be an obvious mistake to suppose that believing is a relation to sentences, the remaining option is that believing is a relation to propositions. However, Prior argues that propositions are 'logical constructions':

> 'Propositions are logical constructions' was first said as a summing-up of this theory [*viz.* Russell's multiple relation theory of judgement]. It meant that statements which appear to be about people and propositions are really about people and quite other things, so that it is not necessary to suppose that there really are such things as propositions.
>
> [Prior, 1971, 8]

Although Prior did not follow Ramsey in adopting a weakened version of Russell's doctrine of real propositional constituents, he did put forward a theory of belief statements that would have the like consequence that attributions of belief are not about propositions.[172] Adopting such a position renders the strong Ramseyan version of deflationism, the version that claims that 'p' and 'The proposition that p is true' are *intensionally* equivalent, more plausible than otherwise. The obvious worry about such a suggestion is that while 'Computers are useful' is about computers, ‘The proposition that computers are useful is true’ seems to be about a proposition. But intensionally equivalent sentences should be about the same thing. In response, a deflationist of this type is likely to claim that, despite appearances, attributions of truth are not about propositions. Ramsey and Prior made this view more plausible by arguing that in all other key contexts (such as belief contexts) statements that seem to be about propositions are really about something else. Prior did not stop at propositions: he suggested that facts are logical constructions too, and statements apparently about them are just about the things mentioned in stating them:

---

[171] See also [Schiffer, 1981].

[172] [Prior, 1971, ch. 2].

> . . . facts and true propositions alike are mere ‘logical constructions’ . . . and . . . they are the *same* ‘logical constructions’ (to have ‘true propositions’ *and* ‘facts’ is to have *too many* logical constructions).
>
> [Prior, 1971, 5]; his italics

And C. J. F. Williams, who largely followed Prior, adds that it is the appeal to truth which explains the locution ‘. . . corresponds to the facts’, not the other way around.

With both facts and propositions out of the picture, the only option seems to be an extreme redundancy theory according to which ‘true’ is not merely eliminable but not even a predicate at all. Both Prior and Williams focus their attention on the uses of ‘true’ that most troubled Ramsey's account — the cases where the proposition to which we attribute truth is not transparently specified.[173] They claim that in saying, for example, ‘What John believes is true’, we are saying (at least to a first approximation) ‘For some p, John believes that p and p’. In taking this Ramsey-like approach, however, they incur an obligation to explain how the quantifier is to be understood. We have already seen the difficulties in trying to read it objectually. This problem is all the more pressing for the current view, for objectual quantification is standardly taken to imply ontological commitment to the entities quantified over. So, if we suppose for a moment that the variables stand for propositions, allowing objectual quantification would undermine Prior and Williams's attempt to do without propositions. But it would also undermine the claim that truth is not a property. For if objectual quantification over propositions is allowed then we can easily define the property that true propositions share as follows:

> For all propositions x [x is true iff $(\exists p)$ ($x =$ the proposition that p and p)].[174]

The usual account of quantifiers suggests that the only other way to read them is substitutionally. Yet this other way relies on its being the case that ‘For some p, John believes that p and p’ is true if and only if there is a sentence such that if we substitute it for '*p*' and erase the quantifier the result will be a true sentence. And this condition might not be met simply because there may be no sentence of our language that expresses John's true belief. So, Prior and Williams settle for neither the objectual nor the usual substitutional reading of the quantifier. Instead, they convert the substitutional reading to one in which the usual biconditional is replaced with a conditional. In other words, 'For some *p*, John believes that *p* and *p*' is true *if* there is a sentence such that if we substitute it for '*p*' and erase the quantifier the result will be a true sentence. If this is right, then Ramsey's original explanation of these contexts has been vindicated. But it leaves the reading of the existential quantifier looking somewhat mysterious; and in fact it must be treated as an undefined primitive.[175]

---

[173] [Williams, 1976].

[174] We have taken this point and this definition from [Soames, 1999, 48].

In order to make the meaning of the existential quantifier here more transparent, even if it remains undefined, Prior suggested that we stop thinking of the propositional (or sentential) variables in this sentence as if they were nominal variables (i.e. variables that stand in for names). Rather, in the same way that nominal variables have their counterparts in ordinary language in the form of pronouns, propositional variables have their counterparts in expressions that have since been called prosentences. For example, 'he' in (10) receives its reference from the previously occurring name 'Jack'.

10. 'Jack is untied and he is angry.'

Now consider (11).

11. 'He explained that he was in financial straits, said that this is how things were, and that therefore he needed an advance.'

Parallelling the use of 'he', we might see 'this is how things were' in (11) as receiving its reference from the previously occurring sentential clause 'he was in financial straits'.[176] Following this line of thought we can read 'There is some *p* such that John believes that *p* and *p*' as 'John believes that things are a certain way, and thus they are'. In such cases, Prior maintains, we simply "extend the use of the 'thing' quantifiers in a perfectly well-understood way ...".[177] The worry that these quantifiers are not properly understood is thus alleviated by the fact that we use and understand these types of expressions in ordinary talk.

But the worry can be entirely removed only if we can find some more commonplace examples of this type of speech. The leading claim of the prosentential theory of truth is that such examples are right in front of us. Prosententialists claim that 'That is true' and 'It is true' often function as what they term 'prosentences', and point out the many parallels between these expressions and pronouns.[178] For example, following Geach, we can point out two types of occurrence for each proform.[179] Sometimes we use proforms out of laziness, rather than use the original expression again. The use of 'he' in (10) is an example of this. But we also sometimes use proforms in quantification contexts, as 'he' is used in (12):

---

[175] [Williams, 1976, 14f]. Prior outlines the truth-conditions for his quantifiers at [Prior, 1971, 35f].

[176] The example, and Prior's idea that certain expressions of English function as prosentences, are taken from [Wittgenstein, 1953, Part I, §134]. Prior makes these points at [Prior, 1971, 38].

[177] [Prior, 1971, 37].

[178] The prosentential theory was first developed by [Grover, Camp and Belnap, 1975]. It has been endorsed by Brandom [1988] and refined by Grover. See [Grover, 1992] for a collection of her essays on the prosentential theory and a useful introduction.

12. Mary loves someone and he is a lucky man.

The expression 'it is true' can function as a proform both in cases of quantification and of laziness. It occurs as a prosentence of laziness in (13):

13. John believes that Brian has been cuckolded and it is true.

It also occurs as a prosentence of quantification in (15) which we can give as a rough reading, ignoring questions of pragmatics and implicature, of (14):

14. What John believes is true.

15. For some proposition, John believes it is true and it is true.

Like Prior and Williams, prosententialists argue that the propositional variables in these quantification cases should be treated non-objectually and non-nominally.

In treating 'It is true' as a prosentence, prosententialists deny that 'true' is a predicate applied to truth-bearers: it is merely part of an expression that picks up its semantic value from a previously used sentence or sentential clause. Further, the 'it' and 'that' of 'It is true' and 'That is true' are not taken as anaphoric pronouns that refer back to a previously used sentence; rather, the whole expression, e.g. 'That is true', functions as a proform. And although prosententialists do not claim that 'true' is redundant (because anaphoric reference is an important semantic device) they do maintain that sentences containing 'true' are equivalent to sentences that do not, or that contain 'true' only as part of a prosentence.[180] Having dispensed with the idea that 'true' functions as a predicate, they infer that there is no property of truth.[181]

These more sophisticated redundancy theories account for the role of 'true' in a wide range of cases and do so with an impressively parsimonious ontology. They also defuse a number of concerns philosophers have had about deflationary theories in general. For one, it is sometimes thought that deflationary theories are committed to some form of anti-realism by their rejection of facts as an ontological category. As Prior points out, though, no such conclusion follows (at least not in any obvious way). 'That the sun is hot is a fact' means the same as 'The sun is hot'; hence, if the sun's being hot is language- and mind-independent, so is the fact that the sun is hot. Redundancy theorists can even claim to capture the 'correspondence intuition' that our thoughts and statements, when true, correspond to the facts, agreeing with Tarski that '... corresponds to the facts' amounts to no more than '... states that p and p'. As Williams makes particularly clear, the correspondence intuition does not amount to a correspondence theory.[182]

---

[179] [Geach, 1962, 124–43].

[180] This is not quite true. The prosentential account first offered by Grover et al. claims that we also need 'true' as part of sentence modifiers such as 'it-is-not-true-that' or 'it-was-true-that'. While it is an important issue as to whether this move succeeds, we must ignore it here.

[181] Grover struggled with the issue of whether she should maintain that truth is a property mainly because she felt that one could give an extension for 'is true'. [Grover, 1992, chapters 6 and 7].

Further, prosententialism overcomes various problems with Ramsey's redundancy theory. Attributions of truth have a pragmatic function ignored in Ramsey's version. They often serve, as Strawson noticed, as endorsements of claims already made. The prosentential theory makes this pragmatic function the centre of its account of truth without succumbing to the Strawsonian exaggeration of supposing that 'true' has no other role. In fact, the important contexts of generalization that Strawson's performative view was unable to deal with can now be easily handled: unlike earlier redundancy accounts, 'true' is deemed to be, precisely *not* redundant, but rather, as Quine urged, important in enabling us to express certain generalizations. The success of this account in explaining these contexts relies on our pre-existing familiarity with the functioning of proforms in quantificational contexts. But it should be emphasized that this success depends on the possibility of an alternative reading of the quantifiers that is neither objectual nor substitutional.

Importantly, having a workable account of these generalization cases provides deflationary theories with an intriguing response to the arguments of Field and others that truth plays an explanatory role in explanations of success or learning from others. It was unsurprising, then, that both the disquotationalist Stephen Leeds and the prosententialists Grover et al. suggested the response at about the same time.[183] While neither Leeds nor Grover considered this particular example, for ease of exposition suppose we agree to the following generalization:

16. If someone has true beliefs about the location of the library, they are more likely to be able to find it.

The occurrence of 'true' in (16) may look as if it provides grounds for supposing that truth has a causal-explanatory role. But it is not clear that it does so if 'true' appears merely as a device of generalization. And it is plausible that we find ourselves asserting (16) simply because asserting the 'true'-free (17) is impossible since it is infinitely long:

17. If someone believes that the library is to the North and the library is to the North, they are more likely to be able to find it, and if someone believes that the library is behind the Physics building and the library is behind the Physics building, they are more likely to be able to find it ... .

Whether one adopts (as does Leeds) Quine's explanation, or prefers the prosentential account of the use of 'true' in expressing generalizations, one can explain why the desire to say something like (17) results in our saying something that involves the use of the truth-predicate: namely, use of this predicate facilitates the expression of a claim that fundamentally has nothing to do with truth. And, indeed, although neither Grover et al. nor Leeds put the point in quite this way, their discussions of these issues ultimately convinced Field that the appearance of 'true' in many of these explanatory contexts was a result, not of the causal-explanatory power of truth, but merely of the fact that 'true' enables the easy expression of handy generalizations.[184]

---

[182] [Williams, 1976, ch. 5].

[183] [Grover *et al.*, 1975, 112; Leeds, 1978, 120–3].

Despite the virtues of prosententialism and the Prior-Williams account, the implication that 'true' is not a predicate and truth not a property makes the position difficult to accept. As Brandom points out, for example, in denying that 'true' is a predicate, prosententialists block off the possibility of treating 'The last thing Bismarck said is true' as a case of predication. Yet in such cases, a disquotational account 'has no greater ontological commitments and stays closer to the apparent form of such sentences'.[185] The apparent logical structure of a range of our inferences involving 'true' also strongly suggests that it is a predicate. Horwich has often insisted in this context that we infer "from '$x =$ *that* p' and 'x is true' to '*that* p is true', and hence to 'p'."[186] Such inferences are most easily made sense of if we suppose that 'true' is here functioning as a predicate. Moreover, there is a weak conception of properties according to which, if we accept that 'true' is a predicate which determines a set, then truth is a property.[187] The Horwich response appears to have formed the current orthodoxy in deflationist thinking (see §3.8).

3.4 Minimal correspondence: Alston, Mackie, Searle

Around the same time as both Field's correspondence theory and the new redundancy accounts were being proposed, John Mackie was advocating what he called 'the simple theory of truth'. The simple theory was largely an attempt to synthesize redundancy and correspondence theories to create a moderate, almost trivial, position. Twenty years later, in the mid-1990s, both William Alston and John Searle championed views that were very similar.

The minimal correspondence view results from abandoning the idea that correspondence is any sort of mirroring of reality by statements, or that there is some sort of one-to-one correlation between parts of a truth-bearer and parts of reality. In developing his own view through a consideration of Austin's correspondence theory, Mackie observed that the latter avoids this mistake, but instead falls into the implication that truth is a 'wholly non-linguistic, non-semantic relation, a situation's being of a certain type'.[188] The only way Mackie could see to retain the spirit of Austin's theory while avoiding this consequence was to suppose that 'to say ... that [a] statement is true is not merely to say that X is of type Y, but to say that *as was stated* X is of type Y.'[189] But once generalized beyond Austin's own kinds of examples, this distils into the claim that 'To say that the-statement-that-p is true is to say things are as they are stated to be'.[190] Minimal correspondence theories thus admit truth-bearers and hold that truth is a relation. This makes a contrast with the deflationary views of the 1970s which abandoned the idea that truth is a relation because they abandoned any commitment to a substantial account of at least one of the *relata*, that is, of propositions or facts.

---

[184] See the afterword to ch. 1 in Field 2001.

[185] [Brandom, 1988, 88]. Brandom suggests the adjustments that a prosententialist can make to accept this while holding on to the claim that 'true' is not a predicate; but, as the quotation suggests, it seems more plausible to suppose that at least in some cases 'true' is functioning as a predicate.

[186] [Horwich , 1998a, 125].

[187] Though examples like 'grue' might make one suspicious of this conception.

Nevertheless, to avoid the mistake of reinstating one-to-one correspondence relations, the minimal correspondence theory usually treats propositions in a recognizably deflationary way. Mackie, for example, emphasizes that propositions are not entities that we have nontrivially postulated in order to do explanatory work:

> Statements or propositions are not (as, say, electrons and genes are) entities by the nontrivial postulating of whose existence we can better explain (and perhaps even predict) what goes on and is observed. The words 'statement' and 'proposition' are just terms that enable us to speak generally about what is said, what is believed, what is assertible or believable and so on.
>
> [Mackie, 1973, 21]

Searle says much the same about facts:

> The something that makes it true that the cat is on the mat is just that the cat is on the mat. And so on for any true statement ... *But we still need a general term for all those somethings, for what makes it true that grass is green, that snow is white, that 2 + 2 = 4 and all the rest. 'Fact' has evolved to fill this need.*
>
> [Searle, 1995, 211]; his italics

On the minimal correspondence view both 'fact' and 'proposition' are general terms but they do not pick out a natural kind or entities that can be called on to do explanatory work.[191]

With minimal facts and propositions in place there is nothing blocking the way to treating truth as a relation between these things. The question, of course, is what this relation amounts to. Both Mackie and Alston, brushing aside as mere cavils the doubts about objectual and substitutional quantification canvassed in the previous section, suggest that the right way to characterize the circumstances under which propositions are true is by using substitutional quantification to generalize what, following Horwich, is now often called the equivalence schema:

---

[188] [Mackie, 1973, 48].

[189] loc. cit.

[190] ibid. 49.

[191] Alston allows that there is work to be done on uncovering the structure of propositions, but sees this work as unnecessary for the theory of truth he offers.

$$(ES) \quad \text{It is true } \textit{that } p \text{ if and only if } p.$$[192]

They follow Prior and Williams in arguing that there is no real barrier to understanding this type of quantification or the use of propositional variables. In characterizing truth in this way, Mackie and Alston think that the nature of correspondence is made clear. Because the same propositional variable occurs on both sides of the biconditional, to say that a statement corresponds to the facts is just to say that how things are stated to be is how things are. Mackie says that if there were any further more complicated relations of correspondence (perhaps of the Fieldian kind) between language or the mind and the world, these are yet to be discovered and play no part in our ordinary understanding of 'true'. Alston, however, allows that there may be a deeper nature to the correspondence relation than has so far been uncovered and that if it were found this would be a significant contribution to the theory of truth.

Mackie and Alston agree, though, that it is a virtue of this way of explaining correspondence that it requires a much tighter relation between facts and true propositions than is normally allowed for by correspondence theories. On this view the two are related, not by mere isomorphism or any other partial similarity, rather, they share the same content.[193] 'If the best we could achieve was that our statements should somehow correspond to what is there, we should still be falling short of having things just as we state them to be', says Mackie, echoing Frege's and Bradley's rehearsals of the identity theory's siren song.[194] But although he agrees with much that Prior and Williams say about the similarity in meaning of 'fact' and 'true proposition', he insists that these are two things that can be related and that only one of them, the fact, is in the world. But *how* the two are related is difficult to ascertain. Indeed, Mackie claims that any attempt to spell out the connection would only lead us to think of it as looser than it is.

Searle's elaboration of the correspondence relation is less equivocal. He makes central use of the disquotation, rather than equivalence, schema, agreeing with the disquotationalist that the schema tells us the conditions under which the statement at issue will be true. He goes on to claim, though, that facts are just these truth-conditions' having been satisfied. We thus have conditions associated with a range of statements; some of these conditions are satisfied. We have the word 'true' for this sub-class of statements. But we also need a word that relates them to the satisfied conditions. 'Correspondence' is a useful expression 'just empty enough and vague enough to allow for all the different kinds of ways in which true statements stand in relation to their relevant fact.'[195] Because of this thin, 'trivial', reading of correspondence, Searle is able to maintain that 'both the correspondence theory and the disquotational theory are true, and they are not in conflict'.[196]

---

[192] [Horwich, 1998a, 6].

[193] [Alston, 1996, 38f].

[194] [Mackie, 1973, 57].

[195] [Searle, 1995, 213].

Of course, it is just this congeniality that is likely to arouse suspicion of the minimal correspondence theory: perhaps it has collapsed into triviality or vacuity. But then minimal correspondence theorists, like deflationists and others, are often happy to accept this consequence:

> I have not said positively in the end any more than it seemed obvious that we should say at the start. If this discussion has any merit, it is in the avoidance of traps into which the example of various distinguished thinkers shows it is all too easy to fall. To give a correct account of the central ordinary sense of 'true' is like walking along a very narrow path with an abyss on either side.
>
> [Mackie, 1973, 57]

Thus it seems to its proposers that if the minimal correspondence view has merit, it will lie in its very minimality. But this has left others dissatisfied, still nagged by the suspicion that there must be more to say about truth than this, since the theory now appears to be mere deflationism and not a version of correspondence at all. The minimal correspondence theory has struggled to find the narrow path between the two abysses of triviality on the one hand and a substantial account of facts on the other. It was perhaps only a matter of time, then, before the idea of extra-linguistic entities that make truth-bearers true, or truthmakers, became part of mainstream philosophy again, but with the radical addition that these truthmakers may not after all be fact-like entities. Mackie's simple correspondence theory, however, may well have been the catalyst that allowed this to occur.

3.5 Truthmakers: Lewis, Mellor

The desire to rehabilitate the correspondence theory was only one of many reasons that truthmakers were eventually readmitted.[197] (As we shall see, closely tied to this motive, though the connection was rarely articulated clearly, was the desire to defend metaphysical realism from the sort of attacks levelled by Dummett, Putnam and Rorty.) But it was an important one. For example, Mulligan et al., like Field, condemned the Tarskian approach on the grounds that it failed to elucidate the relations between language and reality and so gave no account of truth.[198] However, the truthmaker approach allowed one to speak of correspondence as the relation of *making true* that holds between truthmakers and truth-bearers without following Field in attempting to elucidate these relations by giving a 'substantial' account of reference. Moreover, as in more traditional approaches to correspondence, the truthmaking approach sees correspondence as an internal relation; it is part of the nature of the truthmaker that it makes certain truth-bearers true. The goal of what we can call the truthmaker project, then, is to give a systematic account of these truthmakers that not only individuates entities capable of performing the truthmaking role, but tells us which truthmakers make which truth-bearers true.[199]

---

[196] loc. cit.

[197] [Mulligan, Simons and Smith, 1984; Fox, 1987] both offer illuminating historical comments about truthmakers. It should be recorded that the resurgence of interest in truthmaking in Australian philosophy, and consequently in the thought of Lewis and Mellor, is at least in part traceable to C. B. Martin's convincing D. M. Armstrong of the slogan 'No truth without a truthmaker'.

[198] [Mulligan *et al.*, 1984, 288f].

Truthmaker theorists, for the most part, share the intuition that truths are true in virtue of something extra-linguistic, often quoting David Lewis's slogan that truth supervenes on being. Read weakly, as the claim that things being as they are entails that some truth-bearers are true and others false, this slogan is harmless. However, the truthmaker project usually has something stronger in mind. Regardless of what truthmakers are taken to be,[200] the project usually relies on some version of what is generally called the Truthmaker Principle.[201] Versions of this principle are stronger and weaker claims to the effect that the truth of truth-bearers is entailed or necessitated by the existence of some *thing*.[202] For example, the strong version of this claim is that every truth requires such a truthmaker. A more moderate version is reminiscent of logical atomism. Only the atomic truths require truthmakers, while the truth of molecular truths is accounted for truth-functionally. A more sophisticated, and more plausible, principle has been suggested by John Bigelow: 'If something is true, then *there must be*, that is to say, there must *exist*, something which makes the actual world different from how it would have been if this had not been true'.[203] Importantly, it is no part of any of these principles that each truth has its own unique truthmaker. Even on the strongest version of the Truthmaker Principle, as long as each truth-bearer has some truthmaker, it is allowable that some truth-bearers share the same truthmaker. This allows the truthmaker project to avoid the worry that truthmakers are the mere shadows of truth-bearers.

If we accept some version of the Truthmaker Principle, it looks as if it is a short step from there to a correspondence theory of truth. On this view the correspondence relation has become the relation or relations of *making true*. In discovering the truthmakers for different truths (and this may well include empirical or scientific discoveries) we can discover the different sorts of relations that hold between truth-bearers and truthmakers and thus gain insight into the nature of truth.[204] In this way, the principle promises to save a rather strong version of correspondence. In particular, after Strawson it seemed that the truth of propositions is a trivial matter and that if the correspondence theory is to have a chance it must take sentences as its truth-bearers. Yet the truthmaker project promises to make correspondence plausible even for propositions. Even better, this approach is in no way wedded to the idea that truths somehow mirror or are isomorphic with the things that make them true. All that is required is that, for a truthmaker T and a proposition p, the proposition that T exists entails p.

---

[199] For reasons of space we are forced to ignore an alternative approach to the truthmaker project defended by [Taylor, 1976; 1985; Barwise and Perry, 1983; Forbes, 1986]. In contrast to the metaphysical approach canvassed here, the alternative approach attempts to extend standard semantic theories so that facts are treated as the semantic values of sentences.

[200] These things are not necessarily facts, but often facts are taken as what makes at least some truths true. Other suggested truthmakers include ordinary objects and property tokens (tropes). Armstrong [1997] is a strong supporter of the role of facts or states-of-affairs as truthmakers. Mulligan *et al.* [1984] argue for 'moments' or tropes as truthmakers.

[201] We say 'usually' because Mellor, for example, allows that there can be contingent truthmaking and so would not subscribe to any version of the Truthmaker Principle. [Mellor, 2004, 10f].

[202] We have borrowed this way of introducing the truthmaker project from [Forrest and Khlentzos, 2000].

[203] [Bigelow, 1988, 126].

But the correspondence theory is not so easily rescued from its historical difficulties, for the truthmaker project still has the worry of negative and universal truths. D. H. Mellor avoids these problems by abandoning the idea that truthmakers must necessitate the truth of the truth-bearer, and by following the *Tractatus*, allowing that truth-functional constructions, including true negations, do not require complex truthmakers.[205] But, as Mellor is aware, such a move severs the truthmaker project from the correspondence project.[206] Bigelow's version of the Truthmaker Principle allows for a similar neat answer to the problem of universal truths, but at the same cost. He suggests that 'All ravens are black' is true not because there is some general fact, but because for that statement to be false something that does not exist (i.e. a non-black raven) would have had to exist. The sole alternative account of universal propositions available to a truthmaker theorist would otherwise appear to be Ramsey's: they are not *propositions*, but rules, and hence not truth-apt at all.[207]

The truthmaker project also has some difficulty accounting for necessary truth because everything is such that the proposition that it exists necessitates every necessary truth. To prevent these truths being made true by the entire universe, there will need to be some principle of relevance limiting what can count as truthmakers, e.g. the project might be restricted to contingent truths. However, the problem of relevance arises even for these.[208] In an argument with the same conclusion as the slingshot, Greg Restall has shown that, if we think of entailment in the standard way, we can derive the claim that all truths have the same truthmaker. To avoid this we need to appeal to some other sort of entailment or necessitation, one that can limit necessitation to relevant truthmakers, if we are to have a plausible version of the Truthmaker Principle.

---

[204] [Drew Khlentzos, 2000, 116], following Stephen Read, has suggested a way of getting a correspondence theory from the Truthmaker Principle: The principle in its strong form says that if a truth-bearer is true there is some existing thing that makes it true. If some existing thing makes a truth-bearer true, then the truth-bearer is true. Hence, a truth-bearer is true *if and only if* there is some existing thing that makes it true. That is, there is an internal connection between the truth of the truth-bearer and the existence of the truthmaker.

[205] The contrast between Wittgenstein and Russell is striking, the latter admitting as truthmakers negative facts [1918, 187–90] and universal facts (ibid. 207), but drawing the line at disjunctive facts (ibid. 185).

[206] Mellor would not be worried by such a consequence. He argues, as we shall, that the truthmaker project should not be construed as a contribution to the theory of truth [Mellor, 2004, 2].

[207] [Ramsey, 1929, 237f].

[208] [Restall, 1996].

But why should we believe the principle in the first place? In particular, why should we suppose that there must always be some *thing* that makes truths true? Perhaps truthmaker theorists should reduce the demand for existents by following Denis Robinson's slightly weakened version of Bigelow's expression of the principle:

> ... every contingent truth supervenes on the natures of things, in the sense that it could not have been false without some difference, either in what exists, or in what qualities and relations existing things have.
>
> [Robinson, 2000, 148]

But this version, in eschewing talk of things that make truths true, is acceptable even to a redundancy theorist and so seems to carry no consequences for the theory of truth at all. Furthermore, it is not clear why the instances of the equivalence schema should not provide enough for the theory of truthmakers. After all, they tell us the conditions under which the truth-bearer will be true; and Mellor claims that the equivalence principle is 'the only theory of truth which truthmaker theorists need'.[209] But, he adds, if we use the equivalence schema to guide us to truthmakers themselves then we will be led astray. One instance of the schema will be (20):

20. The proposition that my mirror image is waving is true iff my mirror image is waving.

However, it surely is not the case that what makes the statement true is that there is some thing, my mirror image, that is waving. The right hand side of the biconditional tells us only the truth-conditions of the truth-bearer and mentions nothing about the metaphysical or physical structure of the world at all.[210] But it is the latter with which the truthmaker project, as Mellor conceives it, is concerned.

If instances of the equivalence schema tell us truth-conditions without telling us about truthmakers, then Mellor is right to suggest that the truthmaker project has nothing to do with the theory of truth; and *a fortiori* on the Field-Davidson assumption that a theory of truth is a theory of truth-conditions, for, as can be seen from (20), the theory of truthmaking is separate from, and posterior to, a theory of truth-conditions. Moreover, Lewis has a further argument to the same conclusion, in the form of a dilemma for correspondence theorists concerning their conception of facts.[211] On one horn, facts are conceived of as true propositions. Yet, Lewis argued, such a conception reduces the correspondence relation to something acceptable to a redundancy theorist. On the other horn is facts conceived as states-of-affairs, a conception belonging to the approach that conjoins a correspondence theory with the truthmaker project. However, Lewis pointed out, the Truthmaker Principle has (21) as an instance:

---

[209] [Mellor, 2004, 2].

[210] Though it may misleadingly suggest some such thing. One of us can recall being baffled in a school physics class by being told of a plane mirror, 'The image is as far behind the mirror as the object is in front.' This attribution of (apparent) properties of the imaged object to another entity, the image itself, is very common; it still bedevils discussion of mental imagery, by both psychologists and philosophers.

21. It's true that cats purr iff there exists something such that the existence of that thing implies that cats purr.

If we accept the equivalence schema then this entails (22):

22. Cats purr iff there exists something such that the existence of that thing implies that cats purr.

Yet (22) is not about truth. It is about 'the existential grounding of the purring of cats'. But if (22) is equivalent to (21), Lewis argued, then (21) is, despite appearances, not about truth either. He goes on to say that the appearance of 'true' in the various instances of the Truthmaker Principle like (21) seems to be a result merely of its role as a device of generalization; his argument is accordingly much like that which Grover et al. and Leeds used against the claim that truth plays an explanatory role. This general sort of argument puts great pressure on the claim that the truthmaker project supports a correspondence theory of truth, suggesting rather it is one of telling us what truthmaking is, what kinds of entities truthmakers are, and what propositions need them (as Mellor claims). This is reminiscent of one of Russell's logical construction programmes, that whose aim was to show how the truth of everyday propositions is grounded in the restricted class of entities recognized within the preferred epistemological and metaphysical framework.

But Lewis's argument seems to depend on the claim that the equivalence schema expresses an *intensional* equivalence. Only on this assumption is it obvious that the fact that (22) is not about truth entails that (21) is not about truth either. However, if we accept that the equivalence is intensional, then related arguments can be constructed that threaten to destroy all theories of truth bar the deflationary. For, as Alston and Lewis both point out,[212] if a theory of truth is meant as a conceptual analysis of truth, or is even meant to be an *a priori* thesis about truth, then it runs into trouble in the presence of the *a priori* equivalence schema. This is particularly obvious in the case of epistemic theories such as that which claims that it is *a priori* that a statement is true iff it is ideally justified. The appropriate instance of the equivalence schema entails that (23) is equivalent to (24):

23. The proposition that cats purr is true iff the proposition that cats purr is ideally justified.

---

[211] [Lewis, 2001].

[212] [Lewis, 2001, 275; Alston ,1996, 208–14]. Lewis's argument has since been challenged (see Vision 2003, and especially the incisive David 2004).

24. Cats purr iff the proposition that cats purr is ideally justified.

So, we are forced to conclude that (24) is also *a priori* true, when it plainly is not.[213] The same holds, *mutatis mutandis*, for any *a priori* correspondence theory that takes utterances as truth-bearers, as it is surely false that cats purr only if someone has made an appropriate utterance. The only correspondence theory capable of avoiding this difficulty is one which takes propositions as truth-bearers, but, *if* Lewis's previous argument is correct, such a theory is not one of truth at all.

3.6 Neopragmatism: Davidson, Putnam, Rorty

For pragmatists, investigating the truthmaking relation, whether as an attempt to elucidate truth or as part of some other metaphysical project, is a wrongheaded way of doing philosophy. The hunt for truthmakers seems to them an effort to get a God's eye view, from outside all contingent conceptual schemes, of the relations between language and reality. In the last few decades of the twentieth century a number of philosophers, most notably Donald Davidson, Hilary Putnam and Richard Rorty, joined the pragmatists in expressing suspicion about both appeals to truthmakers and attempts to obtain such a God's eye view. And while none of these philosophers, ultimately, endorsed the Jamesian slogan that truth is what is useful, they shared an approach to, and theses about, the nature of truth that justifies the label 'Neopragmatist'. In particular, each pursued the pragmatist goal of breaking down traditional distinctions which they took to have no pragmatic significance, such as those between subjective and objective, value and fact, analytic and synthetic, conceptual scheme and reality, justification and truth. Their inherited pragmatist distaste for these distinctions led them likewise to attempt to simultaneously dissolve the traditional set of problems about realism and truth. But the neopragmatists (ultimately) tried to achieve this by finding a way *between* correspondence and epistemic theories of truth.[214]

This scepticism about truthmaking and the truth/justification distinction led them to reject a claim that guided much of the thought about truth in the twentieth century, including, curiously enough, that of the pragmatists themselves. The claim is the familiar one that truth is a norm of assertion or a goal of inquiry. Davidson has succinctly expressed the divergence from the pragmatists on this point:

---

[213] In effect, this is a formalization of Frege's and Bradley's transparency argument to the conclusion that any analysis of truth must fail. It is also strongly reminiscent of Russell's objection to pragmatism that it implies that in determining the truth of the belief that God exists we should find out whether it is useful to believe that God exists, rather than whether God exists (see §1.5).

[214] For reasons of space, we shall focus only on views shared by all of them and note the differences of opinion in passing. Also, Davidson and Putnam both changed their minds quite significantly even after they began defending neopragmatism. We have tried to present the most considered versions of their views, and largely ignore those they abandoned over this period.

> From the fact that we will never be able to tell which of our beliefs are true, the pragmatists conclude that we may as well identify our best researched, most successful, beliefs with the true ones, and give up the idea of objectivity ... I agree with the pragmatists that we cannot consistently take truth to be both objective and something to be pursued. But I think they would have done better to cleave to a view that counts truth as objective, but pointless as a goal.
>
> [Davidson, 2000, 67][215])

On this view, although truth may be a central concept in our explanations of others, it plays no role in guiding our actions.[216] As we have seen, however, according to Ramsey's success semantics, truth can be treated as both objective and still a goal, in the sense that the utility of true beliefs means that if one wants a belief about P at all (i.e. 'P' or '$\sim P$'), one will prefer the true one and will assess the methods of getting such beliefs accordingly.

Despite this divergence, neopragmatists join the original pragmatists in rejecting the common correspondentist claim that epistemically ideal theories might nevertheless be false. Davidson, for example, argues that we need to use a principle of charity in interpreting others which implies that our interpretations must make as many of their beliefs come out true as possible. Hence, he claims, there is no possibility of any thinker's being *radically* mistaken. Putnam, on the other hand, argued that for any system of belief which is claimed to be radically mistaken there is an equally legitimate, alternative interpretation according to which the beliefs are true. Nor does he allow that we can fix the reference of our terms, and thus ensure a standard interpretation according to which the beliefs are mostly false, by ascending to a meta-language to speak of any causal or referential relations holding between our language and reality. According to Putnam, hoping that we could make such a move is hoping for an ascent to a God's eye view; the reason we cannot so ascend is that our words 'cause' and 'refer' themselves stand in need of interpretation and they too can be provided with an innumerable array of satisfactory interpretations;[217] further, these theories of reference are just more theory requiring an interpretation.

This line of thought lies behind Putnam's famous model-theoretic argument, whose aim was to show the incoherence of even Fieldian fact-free correspondence theories.[218] Putnam relied on the idea of a plurality of reference schemes to conclude that for any language there are innumerable interpretations that will result in the same sentences' being assigned not only the same truth-values, but the same truth-conditions. Adding the assumption that only the truth-conditions of sentences constrain an interpretation of a speaker leads to the conclusion that there is no standard interpretation of a language. Putnam thought that this conclusion is absurd, and that, for reasons that are obscure to us, a semantics based on the idea that truth is ideal justification is the only way to avoid it. Nevertheless, if his basic line of thought is correct, there cannot be a physicalistically respectable account of reference of the sort Field was after because there is no standard reference relation to reduce.[219] The model-theoretic argument, however, is as controversial as the Dummettian anti-realist arguments that inspired it. In particular, Lewis has diagnosed the argument's persuasiveness as depending on overlooking the difference between two different ideas: one, that constraints on interpretations are merely more theory, itself requiring interpretation; the other, that they are constraints to which any overall interpretation of our language must conform.[220]

---

[215] As Davidson notes, Rorty strangely seems to be suspicious of both the distinction between justification and truth and the claim that truth cannot be a goal of inquiry. However, we think that Rorty distinguishes truth from justification for some purposes, including those Davidson himself points out. See [Davidson, 2000, 74, fn. 3].

[216] Stephen Stich has offered an argument against the idea that truth is worth striving for that has much in common with the model-theoretic argument discussed below. He argues that if we agree that there is no standard interpretation of our language, and likewise no standard way of attributing content to our mental states, then it is hard to see why we would want belief states that are true according to one interpretation rather than some other interpretation [Stich, 1990, ch. 5].

[217] [Davidson, 1979, 234f; Putnam, 1978, 126].

However, even if this argument fails, the neopragmatists have a more fundamental reason for rejecting Fieldian correspondence theories – their meaning holism. For example, Davidson's principle of charity is clearly a holistic constraint on interpretation, requiring an interpreter to consider the whole range of beliefs being attributed to a speaker in attributing any particular belief to them.[221] While Putnam and Rorty deny that formal, Davidsonian theories of meaning are possible, they are also committed to meaning holism based on a holism about interpretation.[222] This method of interpretation allows the extension of the reference relation to fall out in whatever way(s) supports the original attributions of content to sentences. Consequently, the neopragmatists do not suppose that the concept of reference need, or could, be reduced to physicalistically respectable concepts. They see 'reference' as a purely theoretical term that gets all its meaning from the way it contributes to the truth theory.[223]

On noting their arguments against correspondence, it seems natural to suppose that the neopragmatists would hold some epistemic theory of truth. Indeed, despite at one time defending correspondence theories of truth, both Putnam and

---

[218] Putnam's [1978] begins with an extended response to Field, and ends with the argument (pp. 123–38). See also [Putnam, 1980; 1983a, 32–5]. A very similar argument can be found in Davidson 1979. Putnam is attacking what he calls 'metaphysical realism', a combination of metaphysical and semantic doctrines, but any anti-realist conclusion comes via the following undermining of the Fieldian notion of correspondence.

[219] Neopragmatists also claim that this argument undermines realism, but Stephen Leeds [1978] shows how one could combine the neopragmatist view of reference schemes with realism.

[220] [Lewis, 1984]. For further criticism on behalf of the realist see [Devitt, 1991]. On behalf of the anti-realist see [van Fraassen, 1997]. Hale and Wright [1997] provide helpful discussion.

[221] [Davidson, 1984].

[222] Putnam 1983b: xvii. Rorty clearly expresses his misgivings about Davidsonian talk of truth-conditions in [Rorty, 2000].

[223] Davidson's [1977] spells this out. The first three chapters of [Putnam, 1978] echo the Davidsonian response to Field, as does [McDowell, 1978].

Davidson have professed allegiance to epistemic theories,[224] while Rorty once endorsed a Peircean notion of truth. In fact, in the mid 1980s it seemed to many that the three of them were united by an anti-realist metaphysics and an epistemic theory of truth. Yet, later, all three either recanted or clarified their position to avoid such attributions.[225] In fact, the neopragmatists have suggested numerous reasons for rejecting epistemic theories. One such reason is that if truth were an epistemic notion then truth could be lost. In other words, because sentences or propositions can be justified at one time and unjustified at another, any theory that built justification into its definition of truth would allow that a truth-bearer could change in truth-value without changing in meaning. This flies in the face of another alleged platitude about truth, namely that truth is stable. Such an argument can be countered, however, by moving, as Putnam did, to a notion of truth as ideal justification.[226] A more serious argument flows from the simple and intuitive claim that statements can be justified and yet not true, i.e., that truth is recognition-transcendent. Putnam's acknowledgement of this point has led to his abandonment of epistemic views and a return to 'common-sense realism'.[227] Yet, as Wright has argued, it is not clear that one needs to give up an epistemic view of truth to agree that truth can be recognition-transcendent.[228]

It seems that there are in fact two basic reasons that the neopragmatists have for rejecting epistemic views. The first is that they are associated with anti-realism and the neopragmatists set out to dispel the appearance of a binary choice between realism and anti-realism by showing that it is based on a mistaken distinction between conceptual schemes and reality.[229] Like the pragmatists, neopragmatists deny that there is some way in which the world is structured independently of the language we use to describe it. This is not meant to imply that we are free to invent the world as we please. That would be to suppose that we are trapped inside our conceptual scheme with reality forever outside. Against that idea, the neopragmatists suggest that this idealism is a result not of abandoning the distinction between scheme and content, but of merely favouring one side of that distinction.[230] What they aim for is a way between these two views, a sort of direct realism that puts us in touch with reality without any representational proxies or intermediaries such as a conceptual scheme.[231] And they see the theory of truth as a key ground on which to fight this battle. While it is not at all clear that theories of truth are relevant to this debate, it *is* clear that they were taken to be so throughout the century. So, to the extent that the neopragmatists wish to deflate or dissolve this debate, it may well be the right dialectical strategy to undermine traditional conceptions of the problems surrounding truth.

---

[224] Contrast [Davidson, 1969] with [Davidson, 1986], and [Putnam, 1960] with [Putnam, 1983a].

[225] [Davidson, 1990, 302, fn. 40; Putnam, 1994]. Rorty [1982, 165] represents such an endorsement (see Rorty 1986: 328, where he retracts it).

[226] [Putnam, 1983a, 55].

[227] Putnam 1994.

[228] [Wright, 2000].

[229] Davidson was the first of the neopragmatists to attack this distinction [Davidson, 1974].

[230] [Rorty, 1986, 326]. Rorty rightly insists that Davidson avoids any such epistemic conception of truth and endorses this stance himself.

[231] 'In giving up the dualism of scheme and world, we do not give up the world, but re-establish unmediated touch with the familiar objects whose antics make our sentences and opinions true or false' [Davidson, 1974, 198]. Putnam's 'internal realism' is best seen as an attempt to reach a similar position. Cf., e.g., [Putnam, 1996].

The other reason neopragmatists have for rejecting epistemic views is that they endorse the claim that truth cannot be defined. Even as Putnam was putting forward his most epistemic view of truth, he conjoined his remarks with the proviso that 'I am not trying to give a formal definition of truth, but an informal elucidation of the notion.'[232] Furthermore, both he and Davidson have often claimed that Tarski has shown us that we cannot provide a consistent definition of truth within our own language. Indeed, Davidson has famously claimed that it is 'folly' to attempt any such definition, as the history of failures to do so bears witness.[233] Instead, he thought, the best we can do is draw out the relations between truth and other semantic and non-semantic concepts. This is what his theory of interpretation is meant to do. By showing how applying a truth-theory to a community requires attributing beliefs and desires and recognizing causal relations between their beliefs and the world, he hoped to elucidate the role truth plays in our explanatory theories of their behaviour.[234] In this way, Davidson attempted to combine the neopragmatist position with primitivism.[235]

Along with the indefinability thesis, Putnam and Davidson share a deflationary attitude without accepting deflationary theories of truth.[236] Davidson sees these theories as yet another attempt to give a definition of truth. Putnam worries that they are committed to a verificationist theory of understanding that cannot allow for a recognition-transcendent conception of truth. Rorty, on the other hand, is far more sympathetic to deflationary theories, seeing in them the hope of dissolving a range of traditional metaphysical and epistemological debates. However, it is not clear that the disagreement here is substantial. Rorty agrees that truth is a property and he also seems to hold Davidson's view that it plays an important explanatory role in theories of interpretation.[237] As will emerge below, and despite Davidson's worry that deflationists try to define 'true', this means that Rorty too can accept primitivism. As for Putnam, he has conceded that truth is not a substantial property while also implying that truth is so variegated that it will evade definition. It seems, then, that, despite their internal squabbles, the neopragmatists combine a deflationary attitude towards truth with primitivism's central claim of its indefinability. We shall see that they share this resting point with a number of other theorists.

---

[232] [Putnam, 1983a, 57].

[233] This is the gist of [Davidson, 1996], and is partly reflected in its title.

[234] See the last section of [Davidson, 1990] for an elaboration of this point.

[235] [Davidson, 1996, 309 and 320ff].

[236] Davidson and Putnam explicitly criticized deflationism and, in particular, Tarski's view read in a deflationary way. See [Davidson, 1990, section 1; Putnam, 1988, 60-71].

[237] The first few sentences of the introduction to [Rorty, 1982] make it clear that Rorty is happy to treat truth as a property. Although in 1986 he insisted that truth was not an explanatory property, he later retracted this claim, emphasizing that he had never meant to deny that truth played a role in a Davidsonian empirical theory of meaning [Rorty, 1995, 282, fn. 23].

3.7 Functionalism and pluralism: Putnam, Rorty, Wiggins, Wright

Yet there is another strain of thought in the neopragmatists that we have so far not considered. There are several different, and not obviously compatible, conclusions one might draw from the purported failure of attempts to discover the nature of truth. One is that there is no nature that requires discovery (deflationism). Another is that truth is unanalysable (primitivism). A third is that there is no *one* property or concept of truth to discover (pluralism). Putnam's turn to 'common-sense realism', for example, was accompanied by an emphasis on the heterogeneous nature of the range of discourses we go in for (ethical, scientific, mathematical, comic ...). Moreover, Putnam counselled, we should not hope for one 'free-standing' explanation of truth that would tell us once and for all what truth is, what a proposition is and what it is for an assertion to be correct.[238] Rather, there are many ways to make a correct assertion and so many ways to correspond to reality.

Rorty, too, has long — indeed, notoriously — urged the pluralist nature of truth. Crucial to his view that truth is not a goal of inquiry is the claim that there is not one goal (Truth-with-a capital-'T', i.e. correspondence with reality) at which all discourses aim. Rather, each has its own standards of justification and correctness, and its own way of 'fitting' reality.[239] This tolerance of different discourses is the basis for Rorty's notoriety, for it suggests an 'anything goes' relativism about both justification and truth. Indeed, he professes a longing for a society that has given up the obsession with truth.[240] Nevertheless, Rorty is attacking merely what he conceives as bad, old truth-with-a-capital-'T': he does not pine for a discourse free of the predicate 'true', a predicate with both an endorsing and a cautionary use (to warn that even if the proposition that snow is white is justified it might still be that snow is not white), and which does not express a substantial property.[241] Thus perhaps he can still allow, and even encourage, the quest to determine whether snow really is white, a subject which one might be forgiven for supposing is a matter of obsession with philosophers.

David Wiggins has also expressed sympathy with pluralism, suggesting a way of developing Davidson's theory of meaning so that we can obtain some sort of pluralist truth theory.[242] He agreed with Davidson that, because of the way truth is related to fact-stating discourse, there must be more to truth than a deflationary reading of Tarski can offer. Yet he also felt the force of Strawson's attack on the notion of facts. Like Field, Wiggins thus wondered how one could provide the sort of substantial theory of truth that was required. He began his attempt with the Davidsonian claim that a theory of truth for a language will serve as a theory of meaning, provided that it meets an anthropological constraint that it make the best sense of the speakers of the language. Accordingly, he thought that to discover the nature of truth we should consider what features any property must have, if it is to serve as the central notion in a theory of meaning so constrained. Wiggins hoped that he could give a functional analysis of the concept of truth by uncovering the relations between such a notion and the anthropological constraint. Inspired by the Ramsey-Lewis approach to defining theoretical terms, he argued that truth would be whichever property satisfied that functional analysis obtained by considering the role any concept must play if it is to be the central concept in the theory of meaning.[243] Unsurprisingly, we have already met, many times, the essential 'marks of truth' which Wiggins claimed to discover: truth is a norm of assertion, recognition-transcendent, and satisfies the correspondence intuition.[244]

---

[238] [Putnam, 1994, 513ff]. His position shows the influence of [Wittgenstein, 1953].

[239] See, in particular, the introduction to Rorty 1982.

[240] [Rorty, 1995, 277ff].

[241] [Rorty ,1986].

[242] [Wiggins, 1980].

Thus, like Davidson, Wiggins contented himself with elaborating the connections amongst truth and other key semantic concepts. He also agreed that one should not try to step outside our linguistic practices in defining truth. However, Wiggins's development of the Davidsonian project lies in the realization that, despite the holistic nature of interpretation, and even, perhaps, despite the irreducibility of the semantic to the physical, one could use this functional analysis to offer a definition of truth. This may not be pluralistic: the functional analysis may be detailed enough to ensure that there is one unique property that fulfils the job description. But the project is consistent with the thought that a number of properties might fit the bill. If one attempts to apply the Davidsonian analysis to a range of discourses, including those that we might doubt are fact-stating, then the pluralistic conclusion may seem even more likely. Perhaps, as Putnam and Rorty suggest, different discourses relate to reality in different ways and so would each require a different property to fill the truth role. Something much like this view has been championed by Crispin Wright.

Wright was inspired to develop his account of truth by his disenchantment with the available ways of expressing anti-realism about the subject matters of different types of discourse. He thought it obvious that it was a mistake to suppose, for example, that anti-realism about the humorous meant that statements about the comic are not capable of being true or false. Following Dummett, he argued that it is not a matter of whether such statements are capable of being true or false, but whether the point of assigning truth-values would be served by assigning them to these statements. This idea led him to a 'minimal' notion of truth that he considers metaphysically 'light-weight' (Wright's expressions) and consequently able to serve as a notion neutral between realists and anti-realists.[245] Like Wiggins, Wright puts forward a number of 'marks of truth' that serve to define the concept. But he does not find these marks by considering the theory of meaning, but by uncovering platitudes about truth – claims concerning truth that we all know *a*

---

[243] ibid. 203f.

[244] This is not an exhaustive list of Wiggins's marks of truth.

[245] [Wright, 1992, 11f].

priori and thus serve to provide an 'analytical theory of the concept'.[246] Chief among these is the equivalence thesis. In a deflationary spirit, Wright argues that many of Wiggins's 'marks of truth' (including the correspondence platitude and the claim that truth is a distinct norm from justification) can be accounted for by any predicate that satisfies this thesis.[247] The platitudes that are uncovered, according to Wright, serve to define a unique concept. Furthermore, he claims, this concept is so light-weight that any assertoric discourse at all, any discourse with an appropriate syntax and constraints on assertibility, will be truth-apt. That is, its declarative sentences will be appropriately assigned truth-values.

One might think that in attempting to show that such a wide range of discourses are truth-apt, Wright was hoping to support the pragmatist cause of dissolving the debate between realism and anti-realism. However, he had quite the opposite goal in mind. Again following Dummett, he argued that this debate is actually a series of local skirmishes and should be seen as a matter of determining what sort of truth-predicate is appropriate for a particular discourse. Thus, although Wright maintains that he can uncover the essential nature of the concept of truth for all discourses, he insists that there may be different properties satisfying the platitudes for different discourses. In particular, he argues that *superassertibility* can play the truth role he has demarcated:

> A statement is superassertible, then, if and only if it is, or can be, warranted and some warrant for it would survive arbitrarily close scrutiny of its pedigree and arbitrarily extensive increments to or other forms of improvement of our information.
>
> [Wright, 1992, 48]

Given that superassertibility is clearly an epistemic property, he concludes that this can serve as the anti-realist's notion of truth. Moreover, the burden is now on the realist to show that in some discourse the property of truth goes beyond superassertibility, and is in no way evidentially constrained.

Wright is explicit that what accounts for the metaphysically light-weight nature of truth is not that he relies only on a set of *platitudes* about truth, but that the notion can be defined for any assertoric discourse. Yet, by his own admission, he offers no argument for the claim that truth can be so defined.[248] But if one agrees with Jackson et al. that a minimal theory of truth does not entail a minimal theory of truth-aptness,[249] one is left wondering why we should accept that all these discourses are truth-apt and thus why we should think of the realist/anti-realist debate as one concerning which property is the realizer for the truth-predicate in a discourse. Another cause for hesitation about Wright's project concerns his suggestion that his pluralism is the only means by which one can understand both sides in the realist/anti-realist debate. A plausible alternative, however, is that although there is only one notion and property of truth, which can be functionally defined, different types of statements have different types of truth-conditions. Some of these truth-conditions may be satisfied independently of our existence and constitution, others may not.[250] If this is a plausible alternative way to such understanding, then the necessity of appealing to pluralism evaporates.

---

[246] [Wright, 1999, 228].

[247] [Wright, 1992, 74, fn 2].

[248] ibid. 74.

[249] [Jackson, Oppy and Smith, 1994].

3.8 Contemporary deflationism: Field, Horwich, Kripke, Soames

Despite the name, Wright's minimalism is not helpfully categorized as a deflationary theory of truth. Although it promises to deliver what he called a light-weight notion of truth, one neutral between realism and anti-realism, it allows that truth might be, for some discourses at least, the sort of property correspondence theorists claim it is. Moreover, Wright claims that determining the notion of truth that is appropriate for a certain discourse is crucial to deciding the realist/anti-realist debate for that local discourse. None of this looks compatible with the deflationary theories so far examined. Nor does Wright's theory belong with the paradigm examples of inflationary accounts — correspondence and epistemic theories. Yet, while a number of positions are similar to Wright's in belonging comfortably in neither camp, by the early 1990s, there was a growing consensus that the most important debate in the theory of truth (and content) was the fundamental one between inflationists and deflationists.[251]

The consensus was fuelled, in part, by the powerful defences of deflationary theories offered by Scott Soames, Hartry Field, and Paul Horwich. Their views depart from most earlier deflationary theories in abandoning the claims that 'true' is not a predicate and does not express a property. However, they upheld a number of doctrines which place them squarely in the deflationist tradition. For one, although both deflationists and inflationists can agree that '... is true' plays a certain logical or expressive role, the contemporary deflationist maintains that the *only* reason for having the concept of truth is to allow us to meet this 'logical need', that is, truth attributions are not descriptive.[252] They do not play their important role by attributing some common characteristic to a truth-bearer, but by serving as a device of disquotation. Another central feature of contemporary deflationism is its commitment to the claim that there is nothing general to be said about the property of truth. Rather truth is such that we grasp, in its entirety, what it is for a particular truth-bearer to be true by grasping the relevant instance of the equivalence thesis: a thesis that also serves, in some sense, to give the meaning of '... is true'. Finally, it is also worth emphasizing an important *conception* of deflationary theories. Some have worried (others, like Rorty, have hoped) that they will have startling philosophical consequences, including the idea that there is no meta-theoretical standpoint from which to assess our theories of reality and so to decide important metaphysical questions. Contemporary deflationists, however, standardly see their theories as having no deep consequences for metaphysics or epistemology; these topics must be separated from the task of providing a theory of truth. As Soames plaintively remarks, 'Throughout the history of philosophy, the notion of truth has occupied a corner into which all manner of problems and confusions have been swept.'[253]

---

[250] [Pettit, 1996] suggests this alternative position.

[251] For example, [Boghossian, 1990, 165, fn. 17]: 'Whether truth is robust or deflationary is the biggest decision a theorist of truth must make.' See also [Field, 1994, 107].

[252] This way of putting the point is Horwich's [1998a, 138].

Field largely based his case for a correspondence theory on the, usually implicit, premiss that truth plays a causal-explanatory role. However, the deflationist response developed by Leeds and Grover et al. convinced him that showing that truth played such a role was exceptionally difficult to do.[254] He thus reconceived the debate about truth as a question of whether, for practical or theoretical purposes, we need to appeal to more than a disquotationalist account. In 1986 he tentatively came down on the side of correspondence. In 1994, and on a number of later occasions, he defended disquotationalism against a number of standard objections, but largely left the issue of truth's causal-explanatory role to one side. The reason for this postponement is that, as previously noted (§3.2), Field sees the crucial issue for a theory of truth as being the question of what determines that utterances have the truth-conditions they do. This meant that, for him, the most important issue between inflationists and deflationists is whether any of our causal explanations makes an ineliminable appeal to the truth-conditions of utterances or belief states. But as Field says, it is a 'big job even to state the worry clearly, and a bigger job to answer it; I must save this for another occasion.'[255] So must we.

However, Field's position must be characterized as deflationism about meaning and content, which is *thereby* deflationism about truth. Deflationism about content is stipulated as being the position that truth-conditions do not play a central (and thus causal-explanatory) role in the theory of content. And he holds that the best way of developing a theory in which neither truth, nor reference and satisfaction, plays such a central role, is to provide disquotational theories of all three. At first, he championed what he called 'pure-disquotationalism' about truth.[256] According to this view, for all utterances p, the assertion of 'p is true' is "cognitively equivalent" to the assertion of p. This claim has a number of odd features that the Field of 1972 would have found repugnant. The first is that the truth-predicate applies only to utterances we can understand. For if we cannot understand p then, because of the *cognitive* equivalence, we will not understand 'p is true' either. The second odd feature is also found in Quine's disquotationalism: sentence-tokens have their truth-conditions necessarily.[257] However, Field argued, this second characteristic is actually a virtue. If we want to describe the nature of space, we can use a truth-predicate and say, perhaps, that all the axioms of Euclidean geometry are true of the space we inhabit. In this case, he suggests, we want, not to make a claim dependent on what certain sentences of English mean, but rather to speak directly about reality. And only a notion of truth for which '*p*' and '"*p*" is true' are cognitively equivalent can allow this.[258] We will return to this point below, but it is anyway clear that our ordinary notion of truth does not share these features. Field, aware of this, proposed that we try to do without the ordinary notion until it can be shown that it serves some purpose beyond those which can be served by a purely disquotational substitute.[259]

---

[253] [Soames, 1999, 255].

[254] See the postscripts to chs 1 and 4 in [Field, 2001].

[255] [Field, 1994, 127].

[256] [Field, 1994].

[257] Field clearly commits himself to this consequence of pure disquotationalism in his [1986, 56-9].

Nevertheless, more recently Field has come to accept a more moderate version of disquotationalism he calls 'quasi-disquotationalism', according to which a notion of translation is built into the understanding of truth; he offers the following definition of quasi-disquotational truth:

> If S is translatable as 'p', then, necessarily, (S is true iff p).[260]

However, he now insists that we individuate sentence-tokens not orthographically, but computationally. That is, two sentence-tokens are equivalent for an individual iff they are treated computationally as equivalent by that individual. This means that sentence-tokens are not capable of being shared across individuals, or within the same individual across possible worlds. Field's account can thus now capture the intuition that sentences have truth-conditions contingently. When we want to attribute truth-conditions to a sentence we must consider which of our actual sentences (individuated computationally) is its translation. Thus 'Rabbits are furry' used by a counterpart me (in another possible world) in a radically different fashion from my use of it, might be translated by my computationally-individuated 'Touch-typing is difficult' and so share that sentence's truth-conditions.[261] The account can also allow that, for us, it is indeterminate whether sentences we do not understand are true. This is an improvement on pure disquotationalism's consequence that such sentences are not true. Finally, if one follows Field in offering a linguistic view of meaning attribution, one can convert the above definition into something that looks a lot like a fully general version of the semantic conception of truth:

> For any sentence S (of any linguistic community C, in any possible world u), if S means that p (for C, in world u) then S (as used by C in world u) is true (at world w) if and only if p (at w).[262]

The conversion is made possible by supposing that 'S means that p' is equivalent to 'S can be translated by my actual sentence "p" .'

Field's trajectory in reaching this point highlights the difficulties faced if one endorses a deflationary theory that takes the primary truth-bearers not to be individuated by their meanings. To mimic our ordinary conception of truth we need to build in to the theory the claim that truth is dependent on meaning. Doing this, however, means that one will also need to give a theory of meaning that does not reconstruct the sort of correspondence account he is trying to avoid.

---

[258] ibid. 58f.

[259] Field proposed that we should be "methodological deflationists". [Field, 1994, 119].

[260] [Field, 2001, 151f].

[261] ibid. 158f.

[262] ibid. 165.

Field himself clearly takes it to be not easy to avoid this error. In fact, he has even accused Soames, who takes himself to endorse a deflationary theory of truth with *propositions* as truth-bearers, of succumbing to it.[263] Thus although Soames has not put forward a theory of truth as such, he provides a problematic case for the distinction between inflationary and deflationary theories and has also made a number of suggestions about the form a deflationary theory should take. His alternative model for a plausible (and deflationary) theory of truth is found in Saul Kripke's influential 'Outline of a Theory of Truth'.[264]

Kripke sought to give a formal construction for a truth-predicate that could deal with paradoxical cases like liar sentences as well as 'ungrounded' statements such as 'This sentence is true'. However, motivated by a number of counter-intuitive consequences of Tarski's hierarchical arrangement of truth-predicates, he wanted his construction to be such that the languages involved contain their own single, univocal truth-predicate. Kripke's guiding thought for defining such a predicate was that it should capture the intuitive idea that 'we are entitled to assert (or deny) of any sentence that it is true precisely under the circumstances when we can assert (or deny) the sentence itself'.[265] In order to deal with cases where the sentence itself contains an attribution of truth, he suggests that we can continue to strip away the truth-predicate until we reach a point at which there are no longer any occurrences of 'true' in the sentence. Having ascertained whether we can assert (or deny) this sentence, our intuitive conception of truth allows us to 're-ascend' to assert (or deny) the original sentence containing 'true'. For many sentences, like '"Snow is white" is true', for example, this technique will yield a determinate truth-value. Kripke's intuitive conception therefore provides the truth-predicate with an extension and an anti-extension. However, for other sentences, including but not restricted to paradoxical sentences, this technique yields no determinate truth-value. We simply cannot trace these sentences, which Kripke calls 'ungrounded', back to some 'true'-free sentence for which we can ascertain that we are entitled to assert (or deny) it. So the anti-extension and extension of the truth-predicate do not together contain all sentences of the language. As Kripke points out, this means that he has only given a partial interpretation for the truth-predicate. There are a variety of sentences that fall in neither the extension nor the anti-extension of the truth-predicate and these are deemed neither true nor false.

Regardless of whether such a proposal really does deal adequately with the paradoxes and related phenomena, [266] Soames sees Kripke's proposal as embodying an important insight: that the problematic cases must be dealt with on the basis of a prior characterization of the truth-predicate.[267] This is not the case with other deflationary proposals. As we shall shortly see, for example, Horwich simply excludes, without explanation, the paradoxical instances of the equivalence schema from the set of instances of the schema that constitute our grasp of the truth concept.

---

[263] [Field, 1986, 66].

[264] [Kripke, 1975].

[265] ibid. 701.

[266] The standard view is that the strengthened liar ('This sentence is untrue') defeats Kripke-style treatments.

So, the seemingly deflationary nature of Soames's views on truth stems from his conception of theories of truth as primarily concerned with giving an account of the meaning of 'true' that (a) does justice to the equivalence between an attribution of truth to a statement and an assertion of the statement itself, and (b) can also deal adequately with the majority of the uses of the predicate, including those within paradoxical contexts. Yet the question remains as to whether he can prevent his conception of truth from inflating under pressure. One source of pressure is that, as Dummett in effect argued, on pain of circularity deflationists who take propositions as truth-bearers are debarred from accepting any account of propositions that individuates them by their truth-conditions. This constraint rules out a wide range of traditional accounts. Further, much that Soames says about the nature of meaning, in particular, suggests that he endorses what Field would call a 'robust' view of what it takes for a sentence to express a proposition.[268] This fact is what led Field to attribute to Soames an inflationary conception of truth. Yet Soames rejects truth-conditional theories of meaning and so rejects the view that we need a robust theory of what it is for sentences to possess their truth-conditions. Thus it is hard to see why his robust theory of meaning should directly entail a robust theory of truth. This is the sort of difficult case which the usual distinctions between inflationary and deflationary theories fail to clearly categorize.

Like Soames, Horwich attempts to avoid disquotationalism's problems by taking propositions as the primary truth-bearers. He also agrees with Soames and Field that he thus needs to give an account of meaning without making essential use of the notion of truth, so that he cannot provide a truth-conditional theory.[269] However, Horwich takes the equivalence schema as the central plank in his minimal theory. More specifically, he claims that minimalism has as its axioms the infinity of instances of that schema. This theory can then be used in his explanation of the meaning of 'true'. He claims that someone grasps the meaning of 'true' when they are disposed to accept all non-paradoxical instances of the equivalence schema without evidence.[270] Thus he does not offer us a definition of 'true' but instead spells out what it is for someone to possess the concept of truth. But he does make a fundamental claim about the property of truth: that there is no more to it than is spelt out by the instances of the equivalence schema.[271] For example, there is no more to the truth of the proposition that snow is white than that snow is white. This is a clear denial of a hidden essence view of truth such as Alston's. Moreover, he insists that all the explanatory work done by appealing to the property of truth can be done using the instances of the equivalence schema. But he draws back from saying that truth is not a property at all; rather, it is not 'a complex or naturalistic property' and so does not have a 'constitutive structure' or a 'causal behaviour' or 'typical manifestations'.[272] He suggests that truth might be best characterized as a 'logical property'.[273]

---

[267] [Soames, 1999, 245f].

[268] See in particular Soames [1984] and [1989].

[269] [Horwich, 1998a, 68ff and 93-103]. However, the reasons Field and Horwich have for agreeing with Dummett on this point seem different. Horwich agrees with Dummett that deflationary theories use the meanings of sentences to define truth and so cannot also use truth to explain meaning. Field, on the other hand, seems motivated by the worry that a truth-conditional theory of meaning will require us to say what makes it the case that sentences have truth-conditions and this will spoil a disquotational theory of possessing truth-conditions. Horwich defends a use theory of meaning in his [1998b].

More radically, Horwich agrees with Field that a deflationist about truth must hold parallel views about reference and satisfaction. The reason for supposing there is such a commitment seems to be that which Field offered in [1972]: namely, that a theory of reference can be used in combination with a Tarski-style truth definition to create a substantial or inflationary theory of truth. Both Field and Horwich also point out that they differ from Tarski in not building compositionality into their definition, focusing instead on the instances of the equivalence schema.[274] As Field says, this frees deflationism from worries about non-compositional natural language constructions by allowing the logic or syntax of expressions to dictate when the axioms about truth will entail compositional theorems. It is clear that the major contemporary forms of deflationism are intended as packages of views about semantics that attempt to deflate the truth-theoretic notions and provide an alternative, non-truth-conditional theory of meaning.

In every such package is the Quinean view that the utility of the truth-predicate lies in its role in allowing us to express certain generalizations. Field, in particular, claimed that the cognitive equivalence of '*p*' and '"*p*" is true' is needed for 'true' to play this role so that even correspondence theorists need a disquotational truth-predicate. Anil Gupta, however, has put serious pressure on this idea, arguing that such a cognitive equivalence would be required only if the generalizations and the conjunction they replace are cognitively equivalent.[275] Yet it is clear that no conjunction, no matter how long and no matter if it exhausts all the instances being generalized over, is cognitively equivalent to a generalization, even if the conjunction, *per impossibile*, is infinitely long. So there is no need for a cognitive equivalence between '*p*' and '"*p*" is true'. However, Gupta's criticism cuts deeper than this. We have also seen that a standard deflationist move is to block assertions of the explanatory power of truth by claiming that when 'true' appears in explanations it does so merely as a device of generalization. Gupta argued that this move assumes that if we can explain all the instances of some generalization we have also explained the generalization. But this in turn relies on the already rejected intensional equivalence.[276] Gupta's point is in fact an expression of the general worry that deflationists will not be able to explain basic generalizations about truth.[277] The problem is particularly worrying for Horwich, who insists that it is only the instances of the equivalence schema that are needed to perform all the explanatory work.[278]

---

[270] [Horwich, 1998a, 35].

[271] ibid. 36f and 135-9.

[272] ibid. 37f.

[273] Horwich takes this term from Field's description of the minimalist position on the nature of truth [Field, 1992, 321].

[274] [Field, 1994, 124f; Horwich 1998a, 10].

[275] [Gupta, 1993, 287-90]. The point is a reprise of [Russell, 1918, 207].

Field offers an obvious and general solution to this worry, one that we have met before: the equivalence schema can be generalized via substitutional quantification and used as the basis for explanations.[279] However, Horwich maintains that the point of the truth-predicate is to enable us to do without substitutional quantification. In other words, if we had a device for substitutional quantification we would not need a truth-predicate at all. Further, Horwich rightly points out that substitutional quantification is usually explained using the notion of truth and so cannot be appealed to in defining it.[280] Although neither point is conclusive, they do suggest that coping with generalizations will not be a trivial matter for a deflationist. Indeed, as Horwich himself observes, there is a basic problem even in giving a general formulation of the deflationist position.

3.9 Primitivism and deflationism: Sosa

Contemporary deflationists usually distinguish themselves from their predecessors by emphasizing that truth is a property. As Ernest Sosa has pointed out, however, this change in the deflationist position brings it very close to primitivism. Field and Horwich both maintain that truth is a property without a constituent structure and which cannot be analysed by providing necessary and sufficient conditions for its instantiation. A Moorean primitivist adheres to the same claims. And although Moore and Russell did not make anything of the equivalence schema, or its instances, as in any way definitional of truth, their version of primitivism looks compatible with its use:

> On this view [Moorean primitivism], what you cannot do either with *good* or with *yellow* or with truth is to define it, to give an illuminating, compact, at least surveyable, Moorean analysis of it. It is in *this* sense that you cannot philosophically "explain" any such "simple" concept. And this leaves it open that you should have *a priori* knowledge of infinitely many propositions constituted essentially by such concepts.
>
> [Sosa, 1993, 11]

---

[276] ibid. 290-5.

[277] This general point is made by Soames [1999, 247f] in criticism of Horwich.

[278] It is not at all worrying for prosententialists, as Gupta notes at fn. 7.

[279] [Field, 1994, 115]. Chris Hill suggests the same move [Hill, 2002].

[280] [Horwich, 1998a, 25f].

Field and Horwich provide explanations of what it is to grasp the concept of truth, rather than merely urging that it is too simple to be analysed, but this does not suffice to distinguish them from primitivists. To suggest what it is to grasp a concept is quite different from offering a definition of the concept. To suppose otherwise would be like thinking one could not be a primitivist about yellow if one claimed that fully possessing the concept of yellow involves being disposed to deploy the concept in the presence of yellow things.

Nor can deflationists readily distinguish themselves from primitivists on the grounds that they point out the supervenience base for the property of truth. While we can agree that they do indicate, for each proposition or utterance, what would make it the case for this truth-bearer to be true, there is nothing in common from one truth-bearer to another summarizable in any sort of analysis or characterization of the property. Primitivists, while not so *committed*, could accept that the property has this massively disjunctive nature. But for some philosophers the possibility of such disjunction counts as a *reductio*.[281] The debate at this point is retracing Moore's waverings of [1899] (see §1.2 above).

As Sosa argued, to go beyond primitivism, deflationists must commit themselves to strong versions of their official claims. For example, if a deflationist said that there is and can be no theory of truth beyond the instances of the equivalence schema, they would be making a claim that is quite foreign to primitivism.[282] The primitivist can accept these biconditionals as *a priori* propositions essentially involving truth, but does not suppose that they are *the only* such propositions. Similarly, primitivists show no signs of agreeing to the standard deflationist claim that truth does not play any explanatory role. However, in these cases it may well seem that primitivism has the upper hand. It is difficult to see why deflationists think these strong conclusions should follow from their basic position, which seems to be just that of the primitivist. Sosa, accordingly, thinks we should adopt the more moderate option and endorse primitivism until the strong deflationary claims can be made out.[283]

These similarities between primitivism and deflationism disturb the generally accepted view that deflationism is incompatible with truth-conditional semantics. In essence, Davidson's position, which is at least *prima facie* consistent, is equivalent to the primitivism of the early Moore and Russell plus the added claim that truth is the central concept in the theory of meaning. Given that the rejection of truth-conditional semantics is meant to follow from basic deflationary claims about truth, however, and given how similar these seem to primitivism, we suggest that the arguments for this incompatibility need to be reconsidered.

---

[281] [Searle, 1995, 215].
[282] [Sosa, 1993, 13f].
[283] ibid. 14.

3.10 Identity: Dodd, Hornsby, McDowell

It is ironic that the twentieth century (and ur-analytic philosophy itself) began with a commitment to a primitivist theory of truth that, as we pointed out in our earlier discussion, is all too easy to confuse with an identity theory. The irony lies in the century's ending with the re-emergence of both primitivism and identity theories which continue to be difficult to keep separate both from each other and from deflationary theories.

The identity theories we have in mind are those officially endorsed by Julian Dodd and Jennifer Hornsby, and strongly suggested by the work of John McDowell. All three are committed to the view that although truth is a property there can be no definition of truth.[284] They also all profess to be attempting to capture the truism that 'When one thinks truly, what one thinks *is* what is the case.'[285] Given this, it does indeed seem as if the identity theory just is primitivism. But the crucial interpretation put on the truism by identity theorists was, of course, that truth-bearers are identical with truthmakers. We have seen the strained metaphysical assumptions which made this interpretation possible for Bradley and his followers. Contemporary identity theorists aim to show us different ways in which the initially shocking identity-statement can be made anodyne.[286]

Dodd's leading idea is that the mistake of correspondence theories is to suppose that there is something that makes a truth true.[287] In other words, correspondentists go wrong in adopting the truthmaker principle. In particular, Dodd has no truck with the idea that the world contains states-of-affairs. To rectify the mistake, he suggests that we do not abandon the intuitive notion of facts, but rather identify facts with true propositions. To avoid letting states-of-affairs creep in through the back door, he maintains that propositions must be conceived as Fregean rather than Russellian. His strategy, then, is to render the identity theory plausible by following Strawson in suggesting that facts are not *in* the world. However, Strawson's further point was that if this is right, then claims about truths corresponding to facts are trivial because there is no way to identify facts independently of identifying the propositions to which they correspond. The same point can be made, and has been, about Dodd's identity theory.[288] In response Dodd has emphasized, as has Hornsby, that it is precisely the point of identity theories that we cannot identify facts independently of propositions.[289] Their positions are maintained in order to point up the failings of the correspondence theory and not to put forward positive claims about the nature of truth. These sound remarkably like the words of a deflationist. Indeed, Dodd claims that, by ruling out truthmakers, arguing for his type of identity theory is precisely the ground-clearing that must be done before one can adopt a deflationary theory — as he went on to do.[290] But his position also remains compatible with primitivism. It is thus difficult to see how his ‘modest identity theory’ is more than deflationism (specifically Horwichian minimalism) plus an added metaphysical claim that states-of-affairs do not exist.

---

[284] [Dodd, 2000, 123; Hornsby, 1997, 2f]. McDowell's notorious quietism suggests that he would agree to the indefinability thesis.

[285] [McDowell, 1994, 27] (his italics). The pull of this idea, and the sense that traditional theories sell us short, is, as we saw in §§1.3 and 3.4, evident in Frege, Bradley and Mackie. Clearly it is not a transitory feature of the notion of truth.

[286] [Candlish, 1999a] explores a range of ways this might be done, some of which resemble the theories of Dodd, Hornsby and McDowell.

[287] [Dodd, 2000, ch. 1].

[288] [Candlish, 1999b, 235f].

[289] [Dodd, 2000, 125; Hornsby, 1997, 3, fn. 5].

Rather than conceive of facts as proposition-like, an alternative way to make an identity theory plausible is to conceive of propositions as fact-like, as the early Russell and Moore did.[291] As Marian David points out in this context, the idea that some propositions are ‘Russellian’ is once again in favour in the wake of the work of Kripke, Putnam, Kaplan and others. Even so, this body of work provides support only for the claim that *some* propositions are Russellian, not all.[292] So, the identity theory of truth for propositions is difficult to defend whether we see propositions as fact-like or facts as proposition-like. And yet, as David also notices, the theory is even more implausible if we treat sentences or belief states as the bearers of truth.[293] Doing so would mean that reality is either linguistic or mental.

Perhaps what is needed to escape these worries is a more radical version of the identity theory like that defended by Hornsby, who takes her lead from some of McDowell's comments in *Mind and World*. Their strategy, in attempting to make the purported identity palatable, is to first introduce the noun ‘thinkable’ as a way of referring to the contents of our propositional attitudes. Like Dodd, though, they agree that thinkables belong in the ‘realm of Fregean sense’.[294] By conceiving of content as that-which-can-be-thought, they make the beginnings of an approach to loosening our grip on the idea that content is radically different from parts of reality. To advance the strategy they ask us to conceive reality as also within the realm of Fregean sense (or within the space of reasons or, again, within the sphere of the conceptual). Like the neopragmatists, they ask us to give up the idea of a reality outside a conceptual scheme. Moreover they do so in order that we can see ourselves as in direct touch with reality without any representational intermediaries.[295] This, they hope, allows us to suppose that there is ‘no ontological gap’ between thought and the world.[296]

Thus, when we think truly we think what is the case because reality itself is thinkable. But why should we suppose this is so? Why agree with the sort of absolute idealism that posits the rationality of reality? Could there not be parts of reality that are unthinkable, that will forever resist our attempts to conceptualize them? There is nothing stopping an identity theorist of this stripe from admitting that not all reality can be conceptualized provided they limit themselves to the claim that *when* a thinkable is true it is identical with an aspect of the world. But McDowell's talk of reality being entirely within the conceptual seems to rule such modesty out. And although Hornsby offers a response to this sort of worry it is difficult to make out what it amounts to.[297] Further, as Dodd points out, the metaphor of the world being within the realm of sense suggests that the world is made up of concepts rather than objects.[298] At the least, the metaphor needs to be explained. Although it is unclear whether this can be done successfully, we think it would be wrong to agree with Dodd that this type of identity theory is obviously internally incoherent. Dodd bases his claim on the assumption that Hornsby and McDowell want facts to be both states-of-affairs and propositions. However, neither part of this assumption is strictly correct. Hornsby has explicitly asserted that facts are within the realm of sense, states-of-affairs are not.[299] On the other hand, 'the realm of Fregean sense' clearly means something quite particular to McDowell. More specifically, it is not meant to be the same 'ontological category' as is opposed to the ontological category of 'the realm of reference' in the traditional dichotomy. Again, though, it is unclear what such a position amounts to and whether it can be defended. The challenge for the identity theorist is the same as that for the minimal correspondentist: to distinguish their view from deflationism without thereby condemning it to obscurity.

---

[290] [Dodd, 2000, chapter 6].

[291] As [Candlish, 1999a] points out, identity theories can result both from nudging facts towards propositions and from nudging propositions towards facts. See also [Engel, 2002, 37-40].

[292] [David, 2001, 692-695].

[293] Ibid: 691f.

[294] Hornsby makes her commitment to this thesis clear in her [1999].

[295] We do not mean to suggest that Hornsby and Dodd are in complete agreement with the neopragmatists on these matters. It is part of McDowell's position, for example, that while Davidson and Rorty share his goals here, they fail, each in their own way, to carry their projects to completion.

[296] [McDowell, 1994, 27]. Hornsby quotes McDowell approvingly at [1997, 1].

4 BRIEF CONCLUSION

Our historical overview of twentieth century analytic philosophy's thinking about truth shows both how diverse this thinking has been and yet how often older theories have been abandoned and forgotten, only to be taken up again. For, on the one hand, theories that share the same name have often proved to be vastly different. Yet, on the other, many of the positions adopted at the close of the century, in an attempt to discover new directions for the philosophy of truth, turned out to be remarkably similar to the theories offered at the beginning of the century which had been largely ignored in the quest for novelty.

Despite the range of different theories of truth that have been, and are still being, endorsed, there appears to be a growing consensus about a number of important claims about truth. The majority of contemporary theorists we have mentioned agree, for example, that there is less to say about truth, and that theories of truth are less central to philosophy, than was once thought. Even those, like Davidson, who see truth as an essential concept for the possibility of thought, do not suppose that the attempt to analyse this concept is a worthwhile philosophical project. While the resulting consensus might be called a deflationary attitude to truth, it should be carefully distinguished from a commitment to what are usually called deflationary theories of truth. The latter are specific theories that embody a number of controversial commitments about truth and related topics, whereas the former is a general feeling that the best we will do when it comes to truth is to spell out a number of truisms and avoid the many mistakes that our brief history has shown it is all too easy to make.

---

[297] [Hornsby, 1997, §§2.3 and 2.4].

[298] [Dodd, 1995, 163]. McDowell attempted to head off this sort of criticism in his [1994, 179f].

[299] [Hornsby, 1999, 241f].

Perhaps part of the explanation for the apparent consensus is that there is a more fundamental, and genuine, consensus that many of the historical debates about truth have been the result either of philosophers focusing their attention unawares on different questions or of disagreements about adjacent philosophical topics for which truth has served as a proxy. We have seen many instances of both types of disagreement. The key issues about truth have been thought, at some periods, to be the use and function of 'true', at others to be the term's definition, at still others to be the criteria for attributing truth to truth-bearers, and at others again to be the task of discovering the nature of the property of truth. Equally fundamental disagreements have arisen because of differences about which truth-bearer is primary and so whether the question of truth is most important for beliefs, or propositions, or utterances or sentences. The confusion has only been compounded by the spilling over, into the theory of truth, of disputes in related areas, such as the theory of meaning, the nature of causal explanations, and the nature of quantification. Often, in fact, the two sources of disagreement we have mentioned came together when divergences about the way to construct theories of meaning, say, resulted in disputes about whether such theories belong to the theory of truth.

As well as being often confused, the extended debate about truth in the previous century also suffered from paying insufficient attention to the nature of truth's apparent opposite, falsity. Although often paid lip service, this topic has rarely been adequately addressed. Yet understanding the relation between falsity, rejection, negation and denial is surely as crucial to understanding our linguistic practice as is understanding the relation between truth, acceptance and assertion. Further, as many philosophers have suggested,[300] theories of truth should allow us to deal with the purported examples of truth-value gaps and gluts. However, understanding falsity and its attendant notions is essential to answering questions about the possibility of such phenomena, and some have argued that the best account of the sense of a sentence lies in giving *both* assertion and denial conditions.[301] Some of the problems informing these debates, in particular the Sorites paradoxes, also suggest that we pay attention to the equally stubborn but neglected question of the possibility of degrees of truth. For one tempting reaction to the Sorites is to postulate continuum-many truth-values between 0 and 1.[302]

---

[300] For example, Kripke and Soames (see §3.8 above).

[301] A rare, but penetrating, discussion of these issues is the debate between Smiley and Priest [1993]. Smiley [1996] revives the occasionally-made case for distinguishing rejection from negation, and Rumfitt [2000] develops Smiley's (p. 4) hint that this may undermine the case for the intuitionist understanding of negation. Dummett, Gibbard and Rumfitt (all 2002) debate the matter.

[302] For a deft description of the reaction and its motivation, some bibliography, and a brief

Apart from these neglected questions, the outstanding issue for the theory of truth is whether it is right to adopt the deflationary attitude. It seems evident that the popularity of this attitude is the result of a failure to reach agreement even on what an investigation of truth ought to be: there is no consensus on how to answer the question, What issues properly belong to the theory of truth?, let alone on whether answering it will involve discovery rather than decision. This does not imply that the deflationary attitude is mistaken: it may be that the reason philosophers have not found a worthwhile project to agree on is that the nature of truth entails that there is no such thing. An alternative explanation is that we have simply lost our way. It remains to be seen which explanation is true.[303]

BIBLIOGRAPHY

In order to avoid misleading the reader about historical matters, the cited date in both the main text and this bibliography is nearly always that of original publication in the original language. The very few exceptions are noted. Where this date differs from that of the *edition* cited, a separate date for the latter is shown.

[Alston, 1996] W. Alston. *A Realist Conception of Truth* (New York: Cornell University Press), 1996.

[Armstrong, 1997] D. Armstrong. *A World of States of Affairs* (Cambridge: Cambridge University Press), 1997.

[Austin, 1950] J. L. Austin. ‘Truth’, *Aristotelian Society Supplementary Volume* 24: 111-29, 1950. Reprinted in S. Blackburn and K. Simmons, eds, 1999: 149–61. Page references are to the latter.

[Austin, 1961] J. L. Austin. ‘Unfair to facts’, in his *Philosophical Papers* (Oxford: Clarendon Press): 102–23, 1961. Reprinted in S. Blackburn and K. Simmons, eds, 1999: 183–200. Page references are to the latter.

[Ayer, 1936] A. J. Ayer. *Language, Truth and Logic*, 1936. Second edition (London: Gollancz, 1946.

[Ayer, 1959] A. J. Ayer. ‘Editor’s introduction,’ in Ayer, A. J. (ed.) *Logical Positivism* (Glencoe, Illinois: The Free Press): 3–28, 1959.

[Baldwin, 1991] T. Baldwin. ‘The identity theory of truth’, *Mind*, 100: 35–52, 1991.

[Baldwin, 1993] T. Baldwin. *G. E. Moore: Selected Writings* (London: Routledge), 1993.

[Barwise and Perry, 1983] J. Barwise and J. Perry. *Situations and Attitudes* (Cambridge, MA: MIT Press), 1983.

[Bigelow, 1988] J. Bigelow. *The Reality of Numbers: A Physicalist’s Philosophy of Mathematics* (Oxford: Clarendon Press), 1988.

[Black, 1948] M. Black. ‘The semantic definition of truth’, *Analysis*, 8: 49-63, 1948.

[Blackburn and SImmons, 1999] S. Blackburn and K. Simmons, eds. *Truth* (Oxford and New York: Oxford University Press), 1999.

[Blanshard, 1939] B. Blanshard. *The Nature of Thought* (London: George Allen and Unwin); 2 volumes, 1939.

[Boghossian, 1990] P. Boghossian. ‘The status of content’, *Philosophical Review*, 99: 157-84, 1990.

---

exploration of its ramifications for, e.g., modus ponens, and assigning truth values to molecular propositions, see chapter 2 of Sainsbury 1995. There is a connection here between the issues surrounding vagueness, and those surrounding paraconsistency: in both cases, the question arises whether the provoking phenomenon (vagueness, inconsistency) belongs merely to the language or can be found in the world itself.

[303] We wish to record our thanks to Hugh Mellor for his valuable comments on an earlier draft.

[Bradley, 1883] F. H. Bradley. *The Principles of Logic*, (London: Oxford University Press, 1883; 1922).

[Bradley, 1893] F. H. Bradley. *Appearance and Reality* (Oxford: Clarendon Press, 1893, 1930).

[Bradley, 1897] F. H. Bradley. Appendix, Index and Explanatory Notes to the second edition of *Appearance and Reality*, 1897.

[Bradley, 1907] F. H. Bradley. 'On truth and copying', *Essays on Truth and Reality* (Oxford: Clarendon Press, 1907, 1914): 107-26. Reprinted in Blackburn, S. and Simmons, K. (eds) 1999. Page references are to the former.

[Bradley, 1909] F. H. Bradley. 'On truth and coherence', *Essays on Truth and Reality* (Oxford: Clarendon Press, 1909, 914): 202-18.

[Brandom, 1988] R. Brandom. 'Pragmatism, phenomenalism, and truth-talk', in French, P., Uehling, T. and Wettstein, H. (eds) *Realism and Antirealism: Midwest Studies in Philosophy*, vol. 12, (Minneapolis: University of Minnesota Press): 75-93, 1988.

[Brandom, 1994] R. Brandom. 'Unsuccessful semantics', *Analysis*, 54: 175-8, 1994.

[Camp, 1975] J. Camp. 'Truth and substitution quantifiers', *Noûs*, 9: 165-85, 1975.

[Candlish, 1989] S. Candlish. 'The truth about F. H. Bradley', *Mind*, 98: 331-48, 1989.

[Candlish, 1996] S. Candlish. 'The unity of the proposition and Russell's theories of judgment', in Monk, R. and Palmer, A. (eds) 1996: 103-33.

[Candlish, 1999a] S. Candlish. 'A prolegomenon to an identity theory of truth', *Philosophy*, 74: 190-220, 1999.

[Candlish, 1999b] S. Candlish. 'Identifying the identity theory of truth', *Proceedings of the Aristotelian Society*, 99: 233-40, 1999.

[Candlish, 2001] S. Candlish. 'Grammar, ontology, and truth in Russell and Bradley', in Gaskin, R. (ed.) 2001: 116-41, 2001.

[Carnap, 1934] S. Carnap. *The Logical Syntax of Language*, first English edition (London: Routledge and Kegan Paul, 1934, 1937).

[Carnap, 1935] R. Carnap. *Philosophy and Logical Syntax* (London: Kegan Paul, Trench, Trubner & Co. Ltd), 1935.

[Carnap, 1942] R. Carnap. *Introduction to Semantics* (Cambridge, MA: Harvard University Press), 1942.

[Carnap, 1949] R. Carnap. 'Truth and confirmation', in Feigl, H. and Sellars, W. (eds) *Readings in Philosophical Analysis* (Atascadero, CA: Ridgeview Publishing Company): 119-27, 1949.

[Cartwright, 1987] R. Cartwright. 'A neglected theory of truth', in his *Philosophical Essays* (Cambridge, MA and London: MIT Press): 71-93, 1987.

[Coffa, 1991] A. Coffa. *The Semantic Tradition From Kant to Carnap* (Cambridge: Cambridge University Press), 1991.

[Daly, 2003] C. Daly. 'Truth and the theory of communication', in Lillehammer, H. and Rodriguez-Pereyra, G. (eds) 2003: 54-67, 2003.

[David, 1994] M. David. *Correspondence and Disquotation* (Oxford: Clarendon Press), 1994.

[David, 2001] M. David. 'Truth as identity and truth as correspondence', in Lynch, M. (ed.) 2001: 683-704.

[David, 2004] M. David. 'Don't forget about the correspondence theory of truth', *Australasian Journal of Philosophy*, 82: 42-7, 2004.

[Davidson, 1967] D. Davidson. 'Truth and meaning', *Synthèse*, 17: 304-23, 1967; reprinted in Davidson 1984: 17-36. Page references are to the latter.

[Davidson, 1969] D. Davidson. 'True to the facts', *Journal of Philosophy*, 66: 748-64, 1969; reprinted in Davidson 1984: 37-54. Page references are to the latter.

[Davidson, 1974] D. Davidson. 'On the very idea of a conceptual scheme', *Proceedings and Addresses of the American Philosophical Association*, 47, 1974; reprinted in Davidson 1984: 183-98. Page references are to the latter.

[Davidson, 1977] D. Davidson. 'Reality without reference', *Dialectica*, 31: 247-53, 1977; reprinted in Davidson 1984: 215-26. Page references are to the latter.

[Davidson, 1979] D. Davidson. 'The inscrutability of reference', *Southwestern Journal of Philosophy*, 10: 7-19, 1979; reprinted in Davidson 1984: 227-41. Page references are to the latter.

[Davidson, 1980] Davidson, D. 1980. *Essays on Actions and Events* (Oxford: Clarendon Press), 1980.

[Davidson, 1984] D. Davidson. *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press).

[Davidson, 1986] D. Davidson. 'A coherence theory of truth and knowledge', in LePore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell): 307-19, 1986.

[Davidson, 1990] D. Davidson. 'The structure and content of truth', *Journal of Philosophy*, 87: 279-328, 1990.

[Davidson, 1996] D. Davidson. 'The folly of trying to define truth', *Journal of Philosophy*, 93: 263-79, 1996; reprinted in Blackburn, S. and Simmons, K. (eds): 308-22. Page references are to the latter.

[Davidson, 2000] D. Davidson. 'Truth rehabilitated', in Brandom, R. (ed.) *Rorty and his Critics* (Oxford: Blackwell): 65-74,2000.

[Devitt, 1991] M. Devitt. *Realism and Truth*, second edition (Princeton, NJ: Princeton University Press), 1991.

[Dewey, 1910] J. Dewey. 'A short catechism concerning truth' in his *The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought* (New York: Peter Smith, 1951): 154-68, 1910.

[Dewey, 1948] J. Dewey. *Reconstruction in Philosophy* (Boston: Beacon Press), 1948.

[Dodd, 1995] J. Dodd. 'McDowell and identity theories of truth', *Analysis*, 55: 160-5, 1995.

[Dodd, 1996] J. Dodd. 'Resurrecting the identity theory of truth: a reply to Candlish', *Bradley Studies,* 2: 42-50, 1996.

[Dodd, 2000] J. Dodd. *An Identity Theory of Truth* (New York: St. Martin's Press),2000.

[Dokic and Engel, 2002] J. Dokic and P. Engel. *Frank Ramsey: Truth and Success* (London: Routledge),2002.

[Dummett, 1959] M. Dummett. 'Truth', 1959, reprinted in Dummett 1978: 1-24.

[Dummett, 1973] M. Dummett. *Frege: Philosophy of Language* (Cambridge, MA: Harvard University Press), 1973.

[Dummett, 1976] M. Dummett. 'What is a theory of meaning?' (II), in Evans, G. and McDowell, J. (eds) 1976: 67-137.

[Dummett, 1978] M. Dummett. *Truth and Other Enigmas* (London: Duckworth), 1978.

[Dummett, 2002] M. Dummett. '"Yes", "no" and "can't say"', *Mind* 111: 289-96, 2002.

[Engel, 2002] P. Engel. *Truth* (Chesham: Acumen), 2002.

[Evans and McDowell, 1976] G. Evans and J. McDowell, (eds). *Truth and Meaning* (Oxford: Clarendon Press), 1976.

[Field, 1972] H. Field. 'Tarski's theory of truth', *Journal of Philosophy*, 69: 347-75, 1972; reprinted in Field 2001: 3-26. Page references are to the latter.

[Field, 1974] H. Field. 'Quine and the correspondence theory', *Philosophical Review*, 83: 200-28, 1974. Reprinted in Field 2001: 199-215. Page references are to the latter.

[Field, 1986] H. Field. 'The deflationary conception of truth', in Macdonald, G. and Wright, C. (eds): 55-117, 1986.

[Field, 1992] H. Field. 'Critical notice: Paul Horwich's *Truth*', *Philosophy of Science*, 59: 321-30, 1992.

[Field, 1994] H. Field. 'Deflationist views of meaning and content', *Mind*, 103: 249-84, 1994. Reprinted in Field 2001: 104-140. Page references are to the latter.

[Field, 2001] H. Field. *Truth and the Absence of Fact* (Oxford: Clarendon Press), 2001.

[Findlay, 1933] J. N. Findlay. *Meinong's Theory of Objects and Values* (Oxford: Clarendon Press, 1933; 1963; revised edition of *Meinong's Theory of Objects*).

[Fogelin, 1987] R. Fogelin. *Wittgenstein* (London: Routledge and Kegan Paul), 1987.

[Forbes, 1986] G. Forbes. 'Truth, correspondence and redundancy', in Macdonald, G. and Wright, C. (eds): 27-54, 1986.

[Forrest and Khlentzos, 2000] P. Forrest and D. Khlentzos. 'Introduction: truth maker and its variants', *Logique et Analyse*, 169-70: 3-15, 2000.

[Fox, 1987] J. Fox. 'Truthmaker', *Australasian Journal of Philosophy*, 65: 188-207, 1987.

[Fraassen, 1997] B. van Fraassen. 'Putnam's paradox: metaphysical realism revamped and evaded', *Philosophical Perspectives*, 11: 17-42, 1997.

[Frege, 1892] G. Frege. 'On concept and object', in *Translations from the Philosophical Writings of Gottlob Frege*, 1892. Translated and edited by Peter Geach and Max Black, second edition (Oxford: Basil Blackwell, 1960): 42-55.

[Frege, 1918] G. Frege. 'Thoughts', 1918. In Frege, *Logical Investigations*, ed. P. T. Geach (Oxford: Blackwell, 1977): 1-30.

[Gaskin, 2001] R. Gaskin, (ed.). *Grammar in Early Twentieth-Century Philosophy* (London and New York: Routledge), 2001.

[Geach, 1962] P. Geach. *Reference and Generality* (Ithaca, N.Y.: Cornell University Press), 1962.

[Gibbard, 2002] P. Gibbard. 'Price and Rumfitt on rejective negation and classical logic', *Mind* 111: 297-304, 2002.

[Grover, 1992] D. Grover. *A Prosentential Theory of Truth* (Princeton, NJ: Princeton University Press), 1992.

[Grover *et al.*, 1975] D. Grover, J. Camp and N. Belnap. 'A prosentential theory of truth', *Philosophical Studies*, 35: 289-297, 1975. Reprinted in Grover 1992: 70-120. Page references are to the latter.

[Gupta, 1993] A. Gupta. 'A critique of deflationism', *Philosophical Topics*, 21: 57-81, 1993. Reprinted in Blackburn, S. and Simmons, K. (eds) 1999: 282-307. Page references are to the latter.

[Haack, 1978] S. Haack. *Philosophy of Logics* (Cambridge: Cambridge University Press), 1978.

[Hale and Wright, 1997] B. Hale and C. Wright. 'Putnam's model-theoretic argument against metaphysical realism', in Hale, B. and Wright, C. (eds) *A Companion to the Philosophy of Language* (Oxford: Blackwell): 427-57, 1997.

[Hempel, 1935] C. G. Hempel. 'On the logical positivists' theory of truth', *Analysis*, 2: 49-59, 1935 and reprinted in Hempel, *Selected Philosophical Essays* ed. by R. Jeffrey (Cambridge: Cambridge University Press, 2000): 9-20. Page references are to the latter.

[Hill, 2002] C. Hill. *Thought and World: an austere portrayal of truth, reference, and semantic correspondence* (Cambridge: Cambridge University Press), 2002.

[Hornsby, 1997] J. Hornsby. Presidential Address: 'Truth: the identity theory', *Proceedings of the Aristotelian Society*, 97: 1-24, 1997; reprinted in Lynch, M. (ed.), 2001. Page references are to the former.

[Hornsby, 1999] J. Hornsby. 'The facts in question: a response to Dodd and to Candlish', *Proceedings of the Aristotelian Society*, 99: 241-5, 1999.

[Horwich, 1994] P. Horwich, (ed.). *Theories of Truth* (Aldershot: Dartmouth), 1994.

[Horwich, 1998a] P. Horwich. *Truth*, second edition, revised (Oxford: Clarendon Press), 1998.

[Horwich, 1998b] P. Horwich. *Meaning* (Oxford: Clarendon Press), 1998.

[Hylton, 1990] P. Hylton. *Russell, Idealism, and the Emergence of Analytic Philosophy* (Oxford: Clarendon Press), 1990.

[Jackson *et al.*, 1994] F. Jackson, G. Oppy and M. Smith. 'Minimalism and truth-aptness', *Mind*, 103: 287-302, 1994.

[James, 1904] W. James. 'Humanism and truth' in James 1909: 37-60, 1904.

[James, 1907a] W. James. 'A word more about truth' in James 1909: 78-89, 1907.

[James, 1907b] W. James. 'Pragmatism's conception of truth' reprinted in Lynch, M. (ed.) 2001: 211-28, 1907.

[James, 1908] W. James. 'The pragmatist account of truth and its misunderstanders' in James 1909: 99-116, 1908.

[James, 1909] W. James. *The Meaning of Truth* (Cambridge, MA and London: Harvard University Press, 1909; 1975)

[Joachim, 1906] H. Joachim. *The Nature of Truth* (London: Oxford University Press, 1906; 1939).

[Khlentzos, 2000] D. Khlentzos. 'What in the world would correspond to truth?', *Logique et Analyse*, 169-70: 109-44, 2000.

[Kirkham, 1992] R. Kirkham. *Theories of Truth: A Critical Introduction* (Cambridge, MA and London: MIT Press), 1992.

[Kneale and Kneale, 1962] W. Kneale and M. Kneale. *The Development of Logic* (Oxford: Clarendon Press), 1962.

[Kripke, 1972] S. Kripke. 'Naming and necessity', in Davidson, D. and Harman, G. (eds) *Semantics of Natural Language* (Dordrecht: D. Reidel): 253-355, 1972.

[Kripke, 1975] S. Kripke. 'Outline of a theory of truth', *Journal of Philosophy*, 72: 690-716, 1975.

[Kripke, 1976] S. Kripke. 'Is there a problem with substitutional quantification?' in Evans, G. and McDowell, J. (eds) 325-419, 1976.

[Leeds, 1978] S. Leeds. 'Theories of reference and truth', *Erkenntnis*, 13: 111-29, 1978.

[Lillehammer and Rodriguez-Pereyra, 2003] H. Lillehammer and G. Rodriguez-Pereyra, eds. *Real Metaphysics: Essays in Honour of D. H. Mellor* (London and New York: Routledge), 2003.

[Lewis, 1984] D. Lewis. 'Putnam's Paradox', *Australasian Journal of Philosophy*, 62: 221-36, 1984.

[Lewis, 2001] D. Lewis. 'Forget about the "correspondence theory of truth"', *Analysis*, 61: 275-80, 2001.

[Lynch, 1997] M. Lynch. 'Minimal realism or realistic minimalism?' *Philosophical Quarterly*, 47: 512-18, 1997.

[Lynch, 2001] M. Lynch, ed.. *The Nature of Truth* (Cambridge, MA and London: MIT Press), 2001.

[Macdonald and Wright, 1986] G. Macdonald. and C. Wright, eds. *Fact, Science, and Morality* (Oxford: Blackwell), 1986.

[Mackie, 1973] J. L.Mackie. *Truth, Probability and Paradox* (Oxford: Clarendon Press), 1973.

[Marcus, 1972] R. Marcus. 'Quantification and ontology', *Noûs*, 6: 240-50, 1972.

[McDowell, 1978] J. McDowell. 'Physicalism and primitive denotation: Field on Tarski', *Erkenntnis*, 13: 131-52, 1978.

[McDowell, 1994] J. McDowell. *Mind and World* (Cambridge, MA: Harvard University Press), 1994.

[Mellor, 2003] D. H. Mellor. '*Real Metaphysics:* Replies', in Lillehammer, H. and Rodriguez-Pereyra, G. (eds), 212-38, 2003.

[Mellor, 2004] D. H. Mellor. 'Truthmakers', unpublished typescript, 2004.

[Monk and Palmer, 1996] R. Monk and A. Palmer, eds. *Bertrand Russell and the Origins of Analytical Philosophy* (Bristol: Thoemmes), 1966.

[Moore, 1899] G. E. Moore. 'The nature of judgment', in Baldwin (ed.) 1993: 1-19, 1899.

[Moore, 1907] G. E. Moore. 'William James' "Pragmatism"', 1907. Reprinted in Moore, *Philosophical Studies* (London: Routledge and Kegan Paul, 1922): 97-146.

[Mulligan *et al.*, 1984] K. Mulligan, P. Simons and B. Smith. 'Truth-makers', *Philosophy and Phenomenological Research*, 44: 287-322, 1984.

[Neale, 1995] S. Neale. 'The philosophical significance of Gödel's slingshot', *Mind*, 104: 761-825, 1995.

[Neurath, 1931] O. Neurath. 'Physicalism,' in Neurath 1983: 52-57, 1931.

[Neurath, 1937] O. Neurath. 'Comments on the papers by Black, Kokoszynska, Williams,' in Neurath 1983: 206-8, 1937.

[Neurath, 1983] O. Neurath. *Philosophical Papers (1913-1946)* edited and translated by Cohen, R. S. and Neurath, M. (Dordrecht, Holland: D. Reidel), 1983.

[Oliver, 1999] A. Oliver. 'A few more remarks on logical form', *Proceedings of the Aristotelian Society*, 99: 247–72, 1999; reprinted (slightly revised) in Gaskin, R. (ed.) 2001: 142-62.

[Palmer, 1988] A. Palmer. *Concept and Object* (London: Routledge), 1988.

[Pears, 1977] D. F. Pears. 'The relation between Wittgenstein's picture theory of propositions and Russell's theories of judgement', *Philosophical Review*, 86: 177-96, 1977.

[Peirce, 1878] C. S. Peirce. 'How to make our ideas clear' in Hartshorne, C. and Weiss, P. (eds) *Collected Papers of Charles Sanders Peirce*, vol. V (Cambridge, MA: The Belknap Press of Harvard University Press, 1878; 1960): 248-71.

[Peirce, 1905] C. S. Peirce. 'What pragmatism is' in Hartshorne, C. and Weiss, P. (eds) *Collected Papers of Charles Sanders Peirce*, vol V (Cambridge, MA: The Belknap Press of Harvard University Press, 1905; 1960): 272-92.

[Pettit, 1996] P. Pettit. 'Realism and truth: a comment on Crispin Wright's *Truth and Objectivity*', *Philosophy and Phenomenological Research*, 56: 883-90, 1996.

[Platts, 1980] M. Platts. 'Introduction', in Platts, M. (ed.) *Reference, Truth and Reality: Essays on the Philosophy of Language* (London: Routledge and Kegan Paul): 1-18, 1980.

[Popper, 1934] K. Popper. *The Logic Of Scientific Discovery*, first English edition (London: Hutchinson, 1934; 1959).

[Popper, 1974] K. Popper. 'Some philosophical comments on Tarski's theory of truth', in Henkin, L. (ed.) *Proceedings of the Tarski Symposium, University of California, June 23-30, 1971* (Providence, RI: American Mathematical Society): 397-408, 1974.

[Priest, 1993] G. Priest. 'Can contradictions be true?', *Aristotelian Society Supplementary Volume*, 67: 35-54 (Part II of a symposium with T. Smiley), 1993.

[Prior, 1971] A. N. Prior. *Objects of Thought* (Oxford: Clarendon Press), 1971.

[Putnam, 1960] H. Putnam. 'Do true assertions correspond to reality?', in his *Mind, Language and Reality: Philosophical Papers Volume 2* (Cambridge: Cambridge University Press, 1975): 70-84, 1960.

[Putnam, 1978] H. Putnam. *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul), 1978.

[Putnam, 1980] H. Putnam. 'Models and reality', *Journal of Symbolic Logic*, 45: 464-82, 1980; reprinted in 1983b: 1-25. Page references are to the latter.

[Putnam, 1983a] H. Putnam. *Reason, Truth and History* (Cambridge: Cambridge University Press), 1983.

[Putnam, 1983b] H. Putnam. *Realism and Reason* (Cambridge: Cambridge University Press), 1983.

[Putnam, 1988] H. Putnam. *Representation and Reality* (Cambridge, MA: MIT Press), 1988.

[Putnam, 1994] H. Putnam. 'Sense, nonsense, and the senses: an inquiry into the powers of the human mind', *Journal of Philosophy*, 91: 488-504 and 510-17, 1994.

[Putnam, 1996] H. Putnam. 'On Wittgenstein's philosophy of mathematics', *Aristotelian Society Supplementary Volume,*70: 243-64, 1996.

[Quine, 1953] W. V. O. Quine. 'Two dogmas of empiricism', in his *From a Logical Point of View*, 1953. (Second edition, revised. Cambridge, MA: Harvard University Press, 1961): 20-46. We have used this version rather than its slightly different ancestor (*Philosophical Review* 1951).

[Quine, 1960] W. V. O. Quine. *Word and Object* (Cambridge, MA: MIT Press), 1960.

[Quine, 1970] W. V. O. Quine. *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice Hall), 1970.

[Quine, 1990] W. V. O. Quine. *Pursuit of Truth* (Cambridge, MA: Harvard University Press), 1990.

[Ramsey, 1923] F. P. Ramsey. 'Critical Notice of *Tractatus Logico-Philosophicus*', in Ramsey 1931: 270-86, 1923.

[Ramsey, 1927] F. P. Ramsey. 'Facts and propositions', in Ramsey 1931: 138-54, 1927. Also in Ramsey 1990. Page references are to the former.

[Ramsey, 1929] F. P. Ramsey. 'General propositions and causality', in Ramsey 1931: 237-55, 1929. Also in Ramsey 1990. Page references are to the former.

[Ramsey, 1931] F. P. Ramsey. *The Foundations of Mathematics* (London: Routledge and Kegan Paul), 1931.

[Ramsey, 1990] F. P. Ramsey. *Philosophical Papers*, ed. Mellor, D. H. (Cambridge: Cambridge University Press), 1990.

[Ramsey, 1927–9] F. P. Ramsey. *On Truth*, ed. Rescher, N. and Majer, U. (Dordrecht: Kluwer), 1927–9; *Episteme*, 16: 1991. (Ch. 1 also in Lynch (ed.) 2001 and in Horwich (ed.) 1994.) [The cited dates are those of likely composition.]

[Restall, 1996] G. Restall. 'Truthmakers, entailment and necessity', *Australasian Journal of Philosophy*, 74: 331-40, 1996.

[Ricketts, 1996] T. Ricketts. 'Carnap: from logical syntax to semantics,' in Giere, R. N. and Richardson, A. W. (eds) *Origins of Logical Empiricism* (Minneapolis: University of Minnesota Press): 231-50, 1996.

[Robinson, 2000] D. Robinson. 'Identities, distinctnesses, truthmakers, and indiscernibility principles', *Logique et Analyse*, 169-70: 145-183, 2000.

[Rorty, 1982] R. Rorty. *Consequences of Pragmatism* (Minneapolis: University of Minnesota Press), 1982.

[Rorty, 1986] R. Rorty. 'Pragmatism, Davidson, and truth' in Lepore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford and New York: Blackwell): 333-55, 1986.

[Rorty, 1995] R. Rorty. 'Is truth a goal of inquiry? Donald Davidson versus Crispin Wright', *Philosophical Quarterly*, 45: 281-300, 1995. Reprinted in Lynch, M. (ed.) 2001: 259-86. Page references are to the latter.

[Rorty, 2000] R. Rorty. 'Response to Donald Davidson', in Brandom, R. (ed.) *Rorty and his Critics* (Oxford: Blackwell): 74-80, 2000.

[Rumfitt, 2000] I. Rumfitt. '"Yes" and "no"', *Mind*, 109: 781-823, 2000.

[Rumfitt, 2002] I. Rumfitt. 'Unilateralism disarmed: a reply to Dummett and Gibbard', *Mind* 111: 305-22, 2002.

[Russell, 1903] B. Russell. *The Principles of Mathematics* (London: Allen and Unwin, 1903; 1937).

[Russell, 1904a] B. Russell. 'Meinong's theory of complexes and assumptions', 1904. In Russell 1994: 432–74.

[Russell, 1904b] B. Russell. 'The axiom of infinity', 1904. In Russell 1994: 476–8.

[Russell, 1905a] B. Russell. 'The existential import of propositions', 1905. In Russell 1994: 486–9.

[Russell, 1905b] B. Russell. 'On denoting', 1905. In Russell 1994: 415–27.

[Russell, 1907] B. Russell. 'The monistic theory of truth', 1907. In Russell 1910a: 131-46.

[Russell, 1908] B. Russell. 'William James's conception of truth', 1908. In Russell 1910a: 112-30.

[Russell, 1910a] B. Russell. *Philosophical Essays.* London: George Allen and Unwin, 1966; originally published with somewhat different contents and different pagination 1910.

[Russell, 1910b] B. Russell. 'On the nature of truth and falsehood', 1910. In Russell 1992: 116–24.

[Russell, 1911] B. Russell. 'Knowledge by acquaintance and knowledge by description',1911. In Russell 1992: 148-61,

[Russell, 1912] B. Russell. *The Problems of Philosophy* (London: Oxford University Press, 1946), 1912.

[Russell, 1918] B. Russell. 'The philosophy of logical atomism', 1918. In Russell 1986: 160–244.

[Russell, 1919] B. Russell. 'On propositions: what they are and how they mean', 1919. In Russell 1986: 278-306.

[Russell, 1924] B. Russell. 'Logical atomism', 1924. In Russell 1988: 160-79.

[Russell, 1959] B. Russell. *My Philosophical Development* (London: George Allen and Unwin), 1959.

[Russell, 1986] B. Russell. *The Collected Papers of Bertrand Russell*, Volume 8*: The Philosophy of Logical Atomism and Other Essays 1914-19* (London: George Allen and Unwin), 1986.

[Russell, 1988] B. Russell. *The Collected Papers of Bertrand Russell, Volume 9: Essays on Language, Mind and Matter 1919-26.* London: Routledge, 1988; 1994 (reprint of Unwin Hyman edition of 1988).

[Russell, 1992] B. Russell. *The Collected Papers of Bertrand Russell*, Volume 6: *Logical and Philosophical Papers 1909-13* (London: Routledge), 1992.

[Russell, 1994] B. Russell. *The Collected Papers of Bertrand Russell*, Volume 4: *Foundations of Logic 1903-05* (London: Routledge), 1994.

[Sainsbury, 1979] R. M. Sainsbury. *Russell* (London: Routledge and Kegan Paul), 1979.

[Sainsbury, 1995] R. M. Sainsbury. *Paradoxes*, second edition (Cambridge: Cambridge University Press), 1995.

[Schiffer, 1981] S. Schiffer. 'Truth and the theory of content' in Parret, H. and Bouveresse, J. (eds) *Meaning and Understanding* (Berlin: W. de Gruyter): 204-22, 1981.

[Schlick, 1934] M. Schlick. 'On the foundation of knowledge', 1934. Reprinted in Schlick 1979: 370-87.

[Schlick, 1935] M. Schlick. 'Facts and propositions,' 1935. Reprinted in Schlick 1979: 400-4.

[Schlick, 1979] M. Schlick. *Philosophical Papers*, ed. Mulder, H. L. and Velde-Schlick, B. F. B. van de (Dordrecht, Holland: D. Reidel), 1979.

[Searle, 1995] J. Searle. *Construction of Social Reality* (New York: Free Press), 1995.

[Smiley, 1982] T. Smiley. 'Presidential Address: The Schematic Fallacy', *Proceedings of the Aristotelian Society*, 83, 1982/83: 1-17, 1982.

[Smiley, 1993] T. Smiley. 'Can contradictions be true?', *Aristotelian Society Supplementary Volume*, 67: 17-33 (Part I of a symposium with G. Priest), 1993.

[Smiley, 1996] T. Smiley. 'Rejection', *Analysis*, 56: 1-9, 1996.

[Soames, 1984] S. Soames. 'What is a theory of truth?', *Journal of Philosophy*, 81: 411-29, 1984. Reprinted in Lynch, M. (ed.) 2001: 397-418. Page references are to the latter.

[Soames, 1989] S. Soames. 'Semantics and semantic competence', *Philosophical Perspectives*, 3: 575-96, 1989.

[Soames, 1999] S. Soames. *Understanding Truth* (Oxford: Oxford University Press), 1999.

[Sosa, 1993] E. Sosa. 'Epistemology, realism, and truth: the first Philosophical Perspectives lecture', *Philosophical Perspectives*, 7, Language and Logic: 1-16, 1993.

[Stalnaker, 1984] R. Stalnaker. *Inquiry* (Cambridge, MA: MIT Press), 1984.

[Stich, 1990] S. Stich. *The Fragmentation of Reason* (Cambridge, MA: MIT Press), 1990.

[Stout, 1907] G. F. Stout. 'Immediacy, mediacy and coherence', *Mind* N.S. 17: 20-47, 1907.

[Strawson, 1949] P. F. Strawson. 'Truth', *Analysis*, 9: 83-97, 1949.

[Strawson, 1950] P. F. Strawson. 'Truth', *Aristotelian Society Supplementary Volume* 24: 129-56, 1950. Reprinted in Blackburn, S. and Simmons, K. (eds) 1999: 162-182. Page references are to the latter.

[Strawson, 1964] P. F. Strawson. 'A problem about truth – a reply to Mr. Warnock', reprinted in his *Logico-Linguistic Papers* (London: Methuen & Co., 1971): 214-33, 1964.

[Tarski, 1933] A. Tarski. 'The concept of truth in formalized languages', in Tarski 1983: 152-278, 1933.

[Tarski, 1936] A. Tarski. 'The establishment of scientific semantics', in Tarski 1983: 401-20, 1936.

[Tarski, 1983] A. Tarski. *Logic, Semantics and Metamathematics* (Indianapolis: Hackett Publishing Company), 1983.

[Taylor, 1976] B. Taylor. 'States of affairs', in Evans, G. and McDowell, J. (eds) 1976: 263-84, 1976.

[Taylor, 1985] B. Taylor. *Modes of Occurrence* (Oxford: Basil Blackwell), 1985.

[Teichmann, 1992] R. Teichmann. 'Whyte on the individuation of desires', *Analysis*, 103-07, 1992.

[Vision, 2003] G. Vision. 'Lest we forget "the correspondence theory of truth"', *Analysis*, 63: 136-42, 2003.

[Walker, 1989] R. C. S. Walker. *The Coherence Theory of Truth* (London: Routledge), 1989.

[Walker, 2001] R. C. S. Walker. 'The coherence theory', in Lynch, M. (ed.): 123-58,, 2001.

[Wallace, 1971] J. Wallace. 'Convention T and substitutional quantification', *Noûs*, 5: 199-211, 1971.

[Warnock, 1964] G. J. Warnock. 'A problem about truth', in Pitcher, G. (ed.) *Truth* (Englewood Cliffs, NJ: Prentice Hall): 54-67, 1964.

[Whyte, 1990] J. T. Whyte. 'Success semantics', *Analysis*, 50: 149-57, 1990.

[Whyte, 1992] J. T. Whyte. 'Weak-kneed desires', *Analysis*, 52: 107-111, 1992.

[Whyte, 1997] J. T. Whyte. 'Success again: replies to Brandom and Godfrey-Smith', *Analysis*, 57: 84-8, 1997.

[Wiggins, 1980] D. Wiggins. 'What would be a substantial theory of truth?' in van Straaten, Z. (ed.) *Philosophical Subjects* (Oxford: Oxford University Press): 189-221, 1980.

[Williams, 1976] C. J. F. Williams. *What is Truth?* (Cambridge: Cambridge University Press), 1976.

[Wittgenstein, 1921] L. Wittgenstein. *Tractatus Logico-Philosophicus*, 1921; first English edition translated by C. K. Ogden (London: Routledge and Kegan Paul, 1922).

[Wittgenstein, 1953] L. Wittgenstein,. *Philosophical Investigations* (Oxford: Blackwell), 1953.

[Wright, 1992] C. Wright. *Truth and Objectivity* (Cambridge, MA: Harvard University Press), 1992.

[Wright, 1995] C. Wright. 'Critical Notice of *The Coherence Theory Of Truth – Realism, Antirealism, Idealism*, by Ralph C. S. Walker', *Synthèse*, 103: 279-302, 1995.

[Wright, 1999] C. Wright. 'Truth: a traditional debate reviewed', *Canadian Journal of Philosophy*, supp. vol. 24: 31-74, 1999. Reprinted in Blackburn, S. and Simmons, K. (eds) 1999: 203-238. Page references to the latter.

[Wright, 2000] C. Wright. 'Truth as sort of epistemic: Putnam's peregrinations', *Journal of Philosophy*, 97: 335-64, 2000.

TRUTH AND PARADOX: A PHILOSOPHICAL SKETCH

JC Beall

1 AIMS AND STRUCTURE OF THIS ESSAY

Because of the many surveys or otherwise readable discussions of truth and paradox, I have tried to take a slightly different (philosophical) perspective on the basic projects and issues. Moreover, I have tried to avoid focusing on the mathematics, leaving citations and ~further reading' to fill out the details; indeed, proofs are entirely left to the cited sources, almost all of which are readily available. The risk in pursuing these aims is that I will inevitably mischaracterize certain approaches, or at least mislead in various ways. That said, the riskiness is probably worth it. And when the risk of distortion or being overly sketchy is high, I try to flag alternative or primary sources to set things straight. (Unfortunately, this has resulted in an enormous amount of footnotes.) This essay is not intended as a representative survey of the many approaches or, per impossibility, issues that are significant in the literature on semantic (or logical) paradox; it aims merely to be a discussion of some of the issues and some of the approaches.

This essay is also abnormal in its chief focus, which is on *dtruth* (see §2), where dtruth is essentially intersubstitutable: dT{A} and A are intersubstitutible in all (non-opaque) contexts,[1] Since the logic of dtruth cannot be classical (assuming the language to be suitably resourceful), theories of dtruth are accordingly nonclassical. I focus largely on a few-broad approaches to dtruth, namely, 'paracomplete' and ~paraconsistent'. On the other hand, many of the most popular approaches to truth-theoretic paradox are not theories of dtruth, but rather theories of some less transparent notion. I sketch but, for space reasons, only very briefly sketch a few such approaches in the last few sections. Further reading is given throughout, and a few remarks for further reading are given in §11. I should also mention that this essay focuses entirely on the semantic side of things. Some sources for axiomatic approaches are indicated in §11.

I have tried to avoid arguing or even commenting much on the adequacy (philosophical or otherwise) of the canvassed proposals. Instead, I try only to mention

---

[1] Note that, for the most part, I trust context to sort out use-mention. Except when it would impede readability, I write (e.g.) T{A} instead of T({A}), despite T's being a unary predicate and, so, standardly written as T(x) or the like. I also leave out discussion of GSdel codes, etc., although this is certainly in the (mathematical) background.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
@ 2007 Elsevier B.V. All rights reserved.

one or two issues that are relevant to the adequacy of the proposals. I leave debate wide open, and otherwise to the cited material.

The basic structure of the essay is as follows. §2 §3 set up the main focus, namely, dtruth and dtruth-theoretic paradox, in particular, Liars. §4 covers a few different 'paracomplete' approaches to dtruth, beginning with a very informal, leisurely discussion of Kripke's non-classical (least fixed point, empty ground model) proposal. §5, in turn, focuses on 'paraconsistent' (and, in particular, 'dialetheic') approaches. §6 closes the discussion of dtruth with a brief discussion of validity and 'dtruth-preservation'. Turning away from dtruth, §7 briefly sketches a few standard parametric and contextualist approaches, including the 'indexiealist' idea, the 'quantifier-variability' proposal, and 'situational truth'. §8, in turn, briefly sketches the basic 'revision-theoretic' idea. The final two sections, §9 and §10, provide a few comments on related matters, with §9 mentioning 'set'-theoretic paradox and its relation to semantics, and §10 (very briefly) discussing 'revenge'. Before turning to the discussion, I should mention the importance of truth-theoretic paradox, something often ignored or, perhaps, unappreciated by many philosophers working on the 'nature' of truth. (It goes without saying that the paradoxes are of fundamental importance in philosophical logic, but they are surprisingly not treated as pressing among philosophers, more generally, even among those who work on 'nature' questions.) By my lights, it is hard to be too invested in theories of the 'nature' of truth without having some grip on, some resolution of, the paradoxes. After all, suppose that one's 'nature' story turns on various key principles. It may well be that such principles cannot consistently or, more broadly, even nontrivially or even plausibly be maintained in a logic that accommodates the paradoxes. (This issue is particularly pressing when it comes to a suitable conditional, a topic discussed below.) Admittedly, the issues of 'nature' and logic are related, since one's 'nature' story will often motivate if not dictate a particular approach to the logic of truth. Still, it strikes me as methodologically misguided to plow" forward on 'nature' questions without at least a steady eye on the truth-theoretic paradoxes. But, for space reasons, I will leave this matter like many in this essay open.

2 DTRUTH AND TRUTH

There is an established usage of 'is true' according to which

(A) is true

and

A

are intersubstitutable in all transparent (non-opaque) contexts. Let us use 'dtrue' for the given usage, a (unary) predicate dT defined simply in terms of such intersubstitutivity.[2] If the language also has a conditional ~ such that every instance of A -+ A is dtrue, then the well-known dT-biconditionals are also, one and all, dtrue: dT(A} +-+A.

We can think of 'dtrue' (or, as I'll say, dtruth) as a mere expressive device, one introduced only to circumvent practical, expressive problems. According to a common metaphor, we once spoke only the 'dtrue'-free fragment of our language. For the most part, the given fragment served our purposes well. We could say that Max is a cat, that GSdel and Tarski were independently ingenious, that there will never be cloned animals, and so on. Daily discourse, so long as it didn't generalize too much, worked well. But generalization is inevitable. Even in daily discourse, let alone science or logic, in general, the need arises to say (what, using dtruth, we say when we say) that all of So-and-so's assertions are dtrue (or that they're false, that is, that the negation of each of So-and-so's assertions is dtrue). Were we God (or even just beings with infinite time or capacities), we wouldn't need to use 'dtrue' in such generalizing contexts; we could simply assert each of So-and-so's assertions (or the negations thereof). But we're not, and so we introduced 'dtrue' to achieve the given sorts of expression. And *that,* and only that, is the job of 'dtrue' in our language.

The foregoing, of course, is but a metaphor, one that is common among so-called deflationists (or, in particular, disquotationalists). Debates rage over whether there is more to *truth* than dtruth, whether there's more to say about the 'nature' of truth than is given in the story about the *device* 'dtrue'. Deflationists, crudely put, maintain that we need not (indeed, ought not) pursue the 'nature' of truth: at bottom, the expressive device *dtr'ue* is all that we need acknowledge,a Non-deflationists maintain that there's more to truth than what is captured by our expressive device *dtrue.*

For present purposes, we can set aside the debate between deflationists and non-deflationists. Both sides can at least acknowledge the simple, 'merely disquotational' usage what we're calling *dtr'uth.* The issue at hand concerns truth-theoretic paradox, which is accentuated in the case of dtruth. Given the intersubstitutivity of dtruth, any distinction between Bivalence (BIV) and the Law of Excluded Middle (LEM) collapses at least with respect to what can be said in the given (object-) language. If, as is standardly assumed, falsity amounts to truth of negation (that A is false exactly if ~A is true), then 'dfalse' and 'not dtrue' are equivalent, that is, dT(~A) and ~dT(A) are equivalent. In turn, dT(A) V ~dT(A) and dT(A} VdT(~A) and A V ~A are equivalent, and so any 'distinction' between LEM and BIV collapses at least in the language. Dtruth is entirely transparent.

---

[2] There is no aim to coin a neologism here, but rather only an attempt to clarify that we're talking about the 'merely expressive' usage of 'true'. I should note that, for purposes of pronunciation, 'dtrue' follows the conventions of Kaplan's 'dthat' [1989, p. 521, fn. 45].

[3] This does not preclude disquotationalists (or deflationists, generally) from acknowledging other 'truth-like' predicates that might do significant work (e.g., in semantics, science). The requirement on disquotationalists is that any other such predicates be ultimately definable in terms of dtruth and other logical devices.

3 LIARS: BROAD PICTURE AND PROJECTS

A typical Liar is a sentence that says or may be used to say of itself only that it is false (or, equivalently in the case of dtruth, not true). Example:

x/ The ticked sentence in §3 is false.

If the ticked sentence is true, it is false, and if false, true. Accordingly, if the ticked sentence is either true or false, it is apparently both true and false. That, in short, is the basic Liar paradox.[4]

Two related but distinct projects dominate the Liar-literature and work on semantic paradoxes, in general, at least among philosophical logicians concerned with modeling truth itself.

NTP. *Non-triviality Project:* Explain how, despite having a truth predicate (in our language, for our language) and Liar-sentences, our language is nontrivial.[5]

ECP. *Exhaustive Characterization Project:*[6] Explain how, if at all, we can truly characterize specify the 'semantic status' of all sentences of our language (in our language).

These projects reflect the core appearances that give rise to the Liar paradox (and its ilk). Semantic paradoxes arise, at least in part, from the appearance that we can 'exhaustively characterize' all sentences of our language in terms of 'semantically significant' predicates, and truly do as much in our language.[7]

Consider the classical picture according to which our semantically significant predicates are 'true' and 'false'. Our exhaustive characterization takes the form of bivalence.

CEC. *Classical:* Every sentence is either true or false.

---

[4] All of this can be formalized precisely. The achievements of Tarski [1936; 1944] (and, relatedly, Gadel) made plain that astonishingly few resources are needed to generate the Liar paradox: a sufficiently 'syntactically rich' language in which the T-biconditionals hold and for which classical logic holds.

[5] *A trivial language* (or theory) is one according to which everything is true. A *nontrivial language* is one that isn't trivial. [For convenience, I will often speak of the (non) triviality of *truth or* 'true' (relative to some language) to mean the same as 'trivial language'. To say that truth or 'true' is trivial is to say that A is true (or A is in the extension of 'true') for *all* A in the given language.[Of course, most theorists are concerned with *consistency* (and, hence, nontriviality); however, 'glut-theorists' and paraconsistentists, in general, are concerned with (reasonable) nontrivial but (negation-) inconsistent languages/theories, and so 'nontriviality' is a more general term. (See §5 for paraconsistent approaches.)

[6] In his [1991], Patrick Grim uses 'complete' in the target sense of 'exhaustive'. Were it not for the already too many notions of 'complete' in logical literature, Grim's terminology would be quite appropriate here. [In particular, one can view the two broad approaches to dtruth, sketched in §4 §5, as in some sense attempting to deal with a choice between (expressive) 'completeness' and (negation) 'consistency'.[

[7] Throughout, I will restrict 'sentence' to declarative (and, for simplicity, context-independent) sentences of a given language. (In some informal examples, context-dependence creeps in, but I trust that no confusion will arise.) Unless otherwise noted, I will leave it implicit that, e.g., 'every sentence' is restricted to the sentences of the given language.

Given such a characterization one that purports to be truly expressible in our language and likewise exhaustive the nontriviality of our truth predicate (and, hence, our language, in general) is immediately in question in the face of Liars the ticked sentence or the like.

The classical picture, of course, is just a special case of the Liar phenomenon. At bottom, there is a tension between the apparent nontriviality of our truth predicate and our language's apparent capacity to achieve (true) exhaustive characterization:

Ec. Exhaustive Characterization: Every sentence is either True, False, or *Other.*

Here, 'Other' is a stand-in for the remaining 'semantically significant predicates',[8] For present purposes, one can focus on the 'problem cases' and think of the semantically significant predicates as those that are invoked to classify such cases (e.g., Liars and the like). For example, if one wishes to classify all Liars (etc.) as *defective in some sense or other,* then 'defective in some sense or other' is semantically significant and thereby stands among one's *Others* in EC.[9]

The Liar paradox makes it difficult to see how we can have both Ec and a nontrivial let alone (negation-) consistent truth predicate. As above, I will (until §7) focus on dtruth (or 'dtrue'), for which intersubstitutivity holds. (See §2.) The question is: How can we, as we appear to, have a nontrivial dtruth predicate (in our language, for our language) and also achieve exhaustive characterization?

In focusing on dtruth, there are two chief approaches: *paracomplete* and *paraconsistent,*[10] These are the approaches that I will discuss until deviating from dtruth at §7 §8.

4 PARACOMPLETE

Typical Liars appear to be sentences that are equivalent to their own negations. This is especially the case with dtruth, where dT(A) and A are intersubstitutable for *all A,* and hence for Liar-like A. The fundamental intersubstitutivity of dT(A) and A is formally modeled by requiring sameness of semantic value: ~(dT(A}) = ~(A) for all A.[11] Where A is a Liar sentence, the requirement is that ~(A) =

---

[8] I will assume throughout that '(d-)true' and '(d-)false' are among our 'semantically significant predicates'. Of course, since I will also assume (throughout) that falsity is truth of negation (i.e., that A is false exactly if ~A is true), one might better put EC just as True(A) V Other{A), with 'Other' as a stand-in, as above. But I will set this aside.

[9] I admit that the notion of 'semantically significant predicate' is not precise, but I think that the target, intuitive sense is clear. Giving a precise account is an open and, in my opinion, quite pressing issue. Unfortunately, except for a few passing comments, I am forced to here leave the notion fairly imprecise, trusting (hoping) that the target idea is clear enough to be useful.

[10] The terminology of 'paracomplete', explained in §4, is from Achille Varzi [1999], also used by Dominic Hyde [1997]. (The related terminology of 'gaps' and 'gluts' was introduced by Kit Fine [1975].)

[11] One can think of *intersubstitutivity* as demanding that A and B entail each other if B differs from A only in containing an occurrence of dT(C} where A contains C. (Here, we assume transparent contexts.)

~(~A), which is impossible in classical semantics. Paracomplete semantics afford more options.

Paracomplete accounts are the dominant approaches towards dtruth. 'Para' in 'paracomplete' comes from the Greek for *beyond* (or, perhaps, *beside).* The classical picture is one according to which every instance of LEM is dtrue, a picture according to which every A or its negation is dtrue. Paracomplete accounts reject LEM and see it as the main principle that paradoxes call into question.[12]

While I will not discuss the issue of *rejection* or *acceptance* in this essay, it is worth noting that in paracomplete (and, indeed, paraconsistent) frameworks, 'reject A' is not usually equivalent to 'accept ~A', at least if, as is the case here, we're dealing with dtruth. If one is rejecting *AV~A* (for some A), then presumably one is thereby or ought, rationally, thereby be rejecting both A and ~A. So, rejecting A is not to be understood as accepting that A is not dtrue, which would be equivalent to accepting ~A. On all of this, paracomplete theorists can (and, presumably, will) agree.[13]

4.1 Kr@ke, Martin Woodruff

The work of Kripke [1975] and Martin Woodruff [1975] was ground-breaking, especially in a climate in which 'Tarskian approaches' were the norm. (See §7.1 for a basic sketch of Tarski's approach.) I will focus on (the nonclassical) interpretation of Kripke's (least fixed point) account.[14] My aim is to give the basic philosophical picture, a sketch of the formal model, and a few comments on apparent virtues and inadequacies of the account. As throughout, I will focus on the semantic picture, tying my remarks to guiding projects NTP and ECP.

Philosophical picture

Recall the dtruth-metaphor according to which 'dtrue' is introduced for purposes of generalization. 'Prior' to introducing the device, we spoke only the 'dtrue'-free fragment. (Similarly for other semantic notions/devices, e.g., 'denotes', 'satisfies', 'true of', etc.) For simplicity, let us assume that the given 'semantic-free' fragment (hence, 'dtrue'-free fragment) is such that LEM holds.[15] Letting £0 be our 'semantic-free fragment', we suppose that A V ~A is dtrue for all A in L20.[16] Indeed, we may suppose that classical semantics and logic, generally is entirely appropriate for the fragment 120-

---

[12] paracomplete accounts are often called 'partiality accounts' or 'partial-predicate accounts'. See [McGee, 19911, [Reinhardt, 1986], and [Soames, 1999].

[13] This will call for a modification of classical probability theory, at least if 'rejection' (similarly, acceptance) is to be understood in terms of probability (e.g., degrees of acceptance being cashed out in terms of probabilities). I will skip discussion here (though see *§4.2* for a brief footnote).

[14] There are various interpretation-issues that surround Kripke's seminal work, both philosophical and logical. See §4.1 for some of the issues.

[15] This assumption sets aside the issue of vagueness (and related sorites puzzles). I am setting this aside only for simplicity. The issue of vagueness or, as some say, 'indeterminacy', in general is quite relevant to some paracomplete approaches to dtruth. See [Field, 2003b], [McGee, 1991], [Soames, 1999].

[16] This assumption is not essential to Kripke's account; however, it makes the basic picture much easier to see.

But now we want our generalization-device. How do we want this to work? As above, we want dT(A) and A to be intersubstitutable for *all A.* The trouble, of course, is that once 'is dtrue' is introduced into the language, various unintended and, given the role of the device, paradoxical sentences emerge (e.g., the ticked sentence above).[17]

The paracomplete idea, of which Kripke's is the best known, is (in effect) to allow some instances of A V ~A to fail.Is In particular, if A itself fails to 'ground out' in 120, fails to 'find a value' by being ultimately equivalent to a sentence in 120, then the A-instance of LEM should fail.[19]

Kripke illustrated the idea in terms of a learning or teaching process. The guiding principle is that dT(A} is to be asserted exactly when A is to be asserted. Consider an 120-sentence that you're prepared to assert say, '1 + 1 = 2' or ~Max is a cat' or whatever. Heeding the guiding principle, you may then assert that '1 + 1 2' and 'Max is a cat' are dtrue. In turn, since you are now prepared to assert

1. 'Max is a cat' is dtrue

the guiding principle instructs that you may also assert

2. "Max is a cat' is dtrue ' is dtrue.

And so on. More generally, your learning can be seen as a process of achieving further and further dtruth-attributions to sentences that 'ground out' in 120. (Similarly for dfalsity, which is just dtruth of negation.) Eventually, your competence reflects precisely the defining intersubstitutivity and transparency of dtruth: that dTIA } and A are intersubstitutable for *all A* of the language.

But your competence also reflects something else: namely, the failure to assert either A or ~A, for some A in the language. To see the point, think of the above process of 'further and further dtruth-attributions' as a process of writing two (very, very big) books one, *The Truth,* the other *The False.* Think of each stage in the process as completing a ~chapter', with chapter zero of each book being empty this indicating that *at the beginning* nothing is explicitly recorded as dtrue (or, derivately, dfalse).2°

Concentrate just on the process of recording *atomics* in *The True.* When you were first learning, you scanned 120 (semantic-free fragment) for the dtrue (atomic)

[17] With respect to formal languages, the inevitability of such sentences is enshrined in G6del's so-called *diagonal lemma.* (Even though the result is itself quite significant, it is standardly called a *lemma* because of its role in establishing G6del Tarski indefinability theorems. For user-friendly discussion of the limitative results, and for primary sources, see [Smullyan, 1992]. For a general discussion of diagonalization, see [Jacquette, 2004].)

lSNB: The sense in which instances of A V ~A 'fail' is modeled by such instances being undesignated (in the formal model). (See §4.1.) How, if at all, such 'failure' is expressed *in* the given language concerns ECP. (See 4.1.)

[19] This is the so-called *least fixed point* picture.

')°Publishers would probably delete chapter zero, but it's worth keeping it for present purposes.

sentences, the sentences you were prepared to assert. Chapter one of *The Truth* comprises the results of your search sentences such as 'Max is a cat' and the like. In other words, letting *'I(t)'* abbreviate *the denotation of t,* chapter one of *The Truth* contains all of those atomics *A(t)* such that *I(t)* exemplifies A, a 'fact' that *would're been* recorded in chapter zero *had* chapter zero recorded the dtrue semantic-free sentences. (For simplicity, if *A(t)* is an /20-atomic such that *I(t)* exemplifies A, then we'll say that *I(t)* exemplifies *A according to chapter" zero.* In the case of 'Max is a cat', chapter zero has it that Max exemplifies cathood, even though neither 'Max is a cat' nor anything else appears in chapter zero.)

In the other book, *The False,* chapter zero is similarly empty; however, like chapter zero of *The Truth,* the sentences that *would* go into *The False's* chapter zero are those (atomic) £0-sentences that, according to the world (as it were), are dfalse e.g., '1 + 1 3', 'Max is a dog', or the like.[21] If A(t) is a dfalse £0-atomic, we'll say that *according to chapter zero, I(t)* exemplifies ~A (even though, as above, chapter zero explicitly records nothing at all). In turn, chapter one of *The False* contains all of those atomics *A(t)* such that, according to chapter zero, *I(t)* exemplifies ~A (i.e., the £0-atomics that are dfalse, even though you wouldn't say as much at this stage).

And now the writing (of atomics) continues: chapter two of *The Truth* comprises 'first-degree dtruth-attributions' and atomics *A(t)* such that, as above, *I(t)* exemplifies A according to chapter one, sentences like (1) and 'Max is a cat'. In turn, chapter three of *The Truth* comprises %econd-degree' attributions, such as (2), and atomics *A(t)* such that (as it were) t is A according to chapter *two.* And so on, and similarly for *The False.* In general, your writing-project exhibits a pattern. Where *I,i(dT)* is chapter i of *The Truth,* the pattern runs thus:

> h+l(dT) = I,~(dT) U {A(t) : *A(t)* is an atomic and *I(t)* exemplifies A according to I,i(dT)}

Let $ comprise all sentences of the language. With respect to *The False* book, the pattern of your writing (with respect to atomics) looks thus:

> I.i+l(dF) I.i(dF) U {A(t) : *A(t)* is an atomic and *I(t)* ~ *8* or *I(t)* exemplifies ~A according to *I,i(dT)}*

So goes the basic process for *atomies.* But what about compound (molecular) sentences? The details are sketched below (see §4.1), but for now" the basic idea is as follows (here skipping the relativizing to chapters). With respect to negations, ~A goes into *The True* just when A goes into *The False.* (Otherwise, neither A nor ~A finds a place in either book.) With respect to *conjunctions, A A t?* goes into *The False* if either A or B goes into *The False,* and it goes into *The True* just if both A and B go into *The True.* (Otherwise, A A B finds a place in neither book.) The case of *disjunctions* is dual, and the quantifiers may be treated as 'generalized conjunction' (universal) and 'generalized disjunction' (existential). 2s

---

[21] For convenience, we'll also put non-sentences into *The False.* Putting non-sentences into *The False* is not essential to Kripke's construction, but it makes things easier. Obviously, one can't *write* a cat but, for present purposes, one can think of *The False* as a special book that comes equipped with attached nets (wherein non-sentences go), a net for each chapter.

Does every sentence eventually find a place in one book or other? No. Consider an atomic sentence L, like the ticked sentence in §3, equivalent to ~dT(L}. In order to get L into *The True* book, there'd have to be some chapter in which it appears. L doesn't appear in chapter zero, since nothing does. Moreover, L doesn't exemplify anything 'according to chapter zero', since chapter zero concerns only the Z20-sentences (and L isn't one of those). What about chapter one? In order for L to appear in chapter one, L would have to be in chapter zero or be such that L exemplifies ~dT(x) according to chapter zero. But for reasons just given, L satisfies neither disjunct, and so doesn't appear in chapter one. The same is evident for chapter two, chapter three, and so on. Moreover, the same reasoning indicates that L doesn't appear in *The False* book.

In general, Liar-like sentences such as the ticked sentence in §3 will find a place in one of our books only if it finds a place in one of the chapters L,(dT) or Li(dF). But the ticked sentence will find a place in I,,(dT) or I,i(dF) only if it finds a place in I.i_l(dT) or I,,_l(dF). But, again, the ticked sentence will find a place in *I.i* l(dT) or *Ii* l(dF) only if it finds a place in I.i 2(dT) or *I.i* 2(dF). And so on. But, then, since Io(dT) and Io(dF) are both empty, and since by our stipulation something exemplifies a property according to Io(dT) only if the property is a non-semantic one (the predicate is in /20), the ticked sentence (or the like) fails to find a place in either book. Such a sentence, according to Kripke, is not only *ungrounded,* since it finds a place in neither book, but also *paradoxical* it *couldn't* find a place in either book. 23

So goes the basic philosophical picture. What was wanted was an account of how, despite the existence of Liars, we could have a dtruth predicate in the language and do so without triviality (or, in Kripke's case, inconsistency). The foregoing picture suggests an answer, at least if we eventually have a chapter I.i(dT) such that dT(A} is in Li(dT) if and only if A is in Li(dT), and similarly a chapter for *The False.* What Kripke (and, independently, Martin Woodruff') showed is that, provided our 'writing process' follows the right sort of scheme (in effect, a logic weaker than classical), our books will contain such target chapters, and in that respect our language can enjoy a (nontrivial, indeed consistent) dtruth predicate. Making the philosophical picture more precise is the job of formal, philosophical modeling, to which I now briefly turn.

Formal model

The main results of [Kripke, 1975] and [Martin and Woodruff, 1975] are more general than I will discuss here. As throughout, my emphasis attempts to be

---

')')This approach to compound sentences reflects the so-called *Strong Kleene* scheme, which is given in §4.1.

2aThe force of *couldn't* here is made precise by the full semantics, but for present purposes one can think of *couldn't* along the lines of *on pain of (negation-) inconsistency* or, for that matter, *on pain of being in both books* (something impossible, on the current framework).

philosophical, with an aim towards ~real truth', with answering NTP and ECP with respect to our 'real language'.

Of course, even though our aim is 'real language', we must none the less abstract a bit from the mess. The aim of formal accounts of truth, at least those concerned with 'real truth' in natural languages (or the very language 'we' speak), is not to give an account of truth, but rather truth-in-Z;, for some formal 'model-language' £. The relevance of such an account is that ~real truth' is supposed to be ~similar enough' to £-truth, at least in relevant respects, to gain answers to NTP and, perhaps, ECP.

For present purposes, I focus on what is known as Kripke's 'least fixed point' model (with empty ground model). I leave proofs to cited works (all of which are readily available), and try to say just enough to see how the formal picture goes.

Following standard practice, we can think of an *interpreted language* £ as a triple (L, A4, ~), where L is the syntax (the relevant syntactical information), M is an ~interpretation' or ~model' that provides interpretations to the nonlogical constants (names, function-symbols, predicates), and cr is a 'semantic scheme' or 'valuation scheme' that, in effect, provides interpretations semantic values to compound sentences.[24]

Consider, for example, familiar classical languages, where the set 12 of 'semantic values' is {1,0}. In classical languages, M = (77,I}, with ~ our (nonempty) domain and I an ~interpretation-function' that assigns to each name an element of ~D (the denotation of the name), assigns to each n-ary function-symbol an element of Z?~ ~ Z?, that is, an n-ary function from Z?~ into ~, and assigns to each n-ary predicate an element of ~D~ ~ 12, a function sometimes thought of as the *intension* of the predicate taking n-tuples of ~ and yielding a ~semantic value' (a ~truth value'). The *eztension* of an n-ary predicate F (intuitively, the set of things of which F is true) contains all n-tuples @1,-.-, an} of D such that *I(F)((at,...,* a~}) 1. The classical valuation scheme ~- (for Tarski) is the familiar one according to which a negation is true (in a given model) exactly when its negatum is false (in the given model), a disjunction is true (in a model) iff one of the disjuncts is true (in the model), and existential sentences are treated as generalized disjunctions.[95]

Classical languages (with suitably resourceful L) cannot have their own dtruth predicate, which is the upshot of Tarski's indefinability theorem. Paracomplete languages reject the 'exhaustive' feature implicit in classical languages: namely, that a sentence or its negation is true, for *all* sentences.

The standard way of formalizing paracomplete languages expands the interpretation of predicates. Recall that in your ~writing process' some sentences (e.g.,

---

[24] For present purposes, a semantic scheme or valuation scheme c~ is simply some general definition of *truth (falsity) in a model.* For a more detailed discussion, see [Gupta and Belnap, 1993].

[25] I assume familiarity with the basic classical picture, including 'true in £' and so on. To make things easier, I will sometimes assume that we've moved to models in which everything in the domain has a name, and otherwise I'll assume familiarity with standard accounts of 'satisfies A(x) in Z2'.

Liars) found a place in neither book. We need to make room for such sentences, and we can expand our semantic values)2 to do so; we can let)2 = {1, ½,0}, letting the middle value represent the status of sentences that found a place in neither book.

Generalizing (but, now, straining) the metaphor, we can think of all n-ary predicates as tied to two such 'big books', one recording the objects of which the predicate is true, the other the objects of which it is false. On this picture, the *eztension* of a predicate F remains as per the classical (containing all n-tuples of which the predicate is true), but we now also acknowledge an *antieztension*, this comprising all n-tuples of which the predicate is false. This broader picture of predicates enjoys the classical picture as a special case: namely, where we stipulate that, for any predicate, the extension and antiextension are jointly exhaustive (the union of the two equals the domain) and, of course, exclusive (the intersection of the two is empty).

Concentrating on the so-called *Strong Kleene* account [1952], 2a the formal story runs as follows. We expand)2, as above, to be {1, s,0}, and so our language £~ (L, jbl, ec} is now a so-called three-valued language (because it uses three semantic values). 2r Our *designated values* intuitively, the values in terms of which *validity* or *eonseq~tenee* is defined are a subset of our semantic values; in the Strong Kleene case, there is exactly one designated element, namely 1.

A (Strong Kleene) model 3// = (/9, I} is much as before, with I doing exactly what it did in the classical case except that I now assigns to n-ary predicates elements of D~ ~ {1, ½, 0}, since)2 {1, ½,0}' Accordingly, the 'intensions' of our paracomplete (Strong Kleene) predicates have three options: 1, ½, and 0. What about *eztensions?* As above, we want to treat predicates not just in terms of extensions (as in the classical languages) but also antiextensions. The *eztension* of an n-ary predicate F, just as before, comprises all n-tuples (al,..., *art}* of ~) such that /(F)((al,..., *a,n})* 1. (Again, intuitively, this remains the set of objects of which F is true.) The *antieztension*, in turn, comprises all n-tuples {al,..., *an}* of D such that *I(F)({al,...,* a~}) 0. (Again, intuitively, this is the set of objects of which F is false.) Of course, as intended, an interpretation might fail to put z in either the extension or antiextension of F. In that case, we say (in our 'metalanguage') that, relative to the model, F is *undefined* for x. 98

26This is one of the paracomplete languages for which Kripke proved his definability result. Martin Woodruff proved a special case of Kripke's general 'fixed point' result, namely, the case for so-called 'maximal fixed points' of the *Weak Kleene* scheme, or *weak Kleene languages.*

2rKripke [1975] made much of emphasizing that 'the third value' is not to be understood as a *third truth value* or anything else other than 'undefined' (along the lines of Kleene's original work [1952]). I will not make much of this here, although what to make of semantic values that appear in one's formal account is an important, philosophical issue, one that I'll briefly touch on in §10. (Note that if one wants to avoid a three-valued language, one can let V = {1, 0} and proceed to construct a Kleene-language by using *partial functions* (hence, the standard terminology 'partial predicates') for interpretations. I think that this is ultimately merely terminological, but I won't dwell on the matter here.

2SA common way of speaking is to say that, for example, *F(t)* is 'gappy' with respect to *I(t).* This terminology is appropriate if one is clear on the relation between one's formal model and the

Letting .7-+ and jr be the extension and antiextension of F, respectively, it is easy to see that, as noted above, classical languages are a special case of (Strong Kleene) paracomplete languages. Paracomplete languages typically eschew inconsistency,[29] and so typically demand that jr+ N jc- 0, in other words, that nothing is in both the extension and antiextension of any predicate. In this way, paracomplete languages (typically) agree with classical languages. The difference, of course, is that paracomplete languages do *not* demand that jr+ U jr :D for all predicates F. But paracomplete languages *allow* for such 'exhaustive constraints', and in that respect can enjoy classical languages as a special case.

To see the close relation between classical languages and Strong Kleene, notice that n, the Strong Kleene valuation-scheme, runs as follows (here treating only 7, V, and B). Where *V~4(A)* is the semantic value of A in A// (and, for simplicity, letting each object in the domain name itself), and, for purposes of specifying scheme ~, treating F as standardly (linearly) ordered:

K1. *V~(~A) 1 VM(A).*

K2. *V~(A* V B) = max(Vjt4(A), VM (B)).

K3. Vj~(~xm(x)) = *max{Vj~(A(t/x)):* fox"all t • ~}.

The extent to which classical logic is an extension of a given paracomplete logic depends on the semantic scheme of the language.[3°] Since n, as above, is entirely in keeping with the classical scheme *7 except* for 'adding an extra possibility', it is clear that every classical interpretation is a Strong Kleene-interpretation (but not *vice versa).*[3x]

Let us say that an interpretation *verifies* a sentence A iff A is designated (in this case, assigned 1) on that interpretation, and that an interpretation verifies a set of sentences ~ iff it verifies every element of ~. We define *semantic consequence* in familiar terms: A is a consequence of N iff every interpretation that verifies also verifies A. I will use 'lb,~' for the Strong Kleene consequence relation, so understood.[32]

Let us say that a sentence A is logically true in £~ exactly if ~ IP~ A, that is, iff A is designated (assigned 1) in every model. A remarkable feature of £~ is that there are no logical truths. To see this, just consider an interpretation that

---

target notions that the model is intended to serve (in one respect or other), but the terminology can also be confusing. I will return to this point below (see §4.2 and §10).

[29] But see §5.2.

[3°] Here, perhaps not altogether appropriately, I am privileging model theory over proof theory, thinking of 'logic' as the semantic consequence relation that falls out of the semantics. This is in keeping with the elementary aims of the essay, even though (admittedly) it blurs over a lot of philosophical and logical issues.

[31] Note that in classical languages, VA4(A) ~ {1, 0} for any A, and the familiar classical clauses on negation, disjunction, etc. are then simply (K1) (K3).

[32] *'1~'* iS increasingly used for *forcing* in set theoretic-logical literature, and so '~' might be a better choice for semantic consequence. The trouble is that the latter is very standardly used in truth-theoretic literature to represent the *true-in-a-language* or *true-in-a-model* relation.

assigns $\overset{1}{\sim}$ to every atomic, in which case, as an induction will show, every sentence is assigned $\overset{1}{\sim}$ on that interpretation. Hence, there's some interpretation in which no sentence is designated, and hence no sentence designated on all interpretations. Afortiori, LEM fails in Strong Kleene languages.[33]

And now an answer to NTP, with respect to dtruth, becomes apparent. What we want is a model of how our language can be nontrivial (indeed, consistent) while containing both a dtruth predicate and Liar-like sentences. In large part, the answer is that our language is (in relevant respects) along Strong Kleene lines, that the logic is weaker than classical logic. Such a language, as Kripke showed, can contain its own dtruth predicate.

The construction runs (in effect) along the lines of the 'big books' picture. For simplicity, let £~ be a *classical* (and, hence, Strong Kleene) language such that L (the basic syntax, etc.) is free of semantic terms but has the resources to describe its given syntax including, among other things, having a name (A) for each sentence. (So, I assigns to each r~-ary predicate an element of ~9~ ~{1, 0}, even though the values F of Z2~ also contain ½.) What we want to do is move to a richer language the syntax Lt of which contains dT(z), a unary predicate (intended to be a dtruth predicate for the enriched language). For simplicity, assume that the domain I) of £~ contains all sentences of Lt.[34]

Think, briefly, about the 'big books' picture. One can think of each successive 'chapter' as a language that expands one's official record of what is dtrue (dfalse). More formally, one can think of each such 'chapter' of both books as the extension and antiextension of 'dtrue', with each such chapter expanding the interpretation of 'dtrue'. Intuitively (with slight qualifications about chapters zero), one can think of ~i+l(dT) as explicitly recording 'what is true' *acco~ding to* chapter I.i(dT). The goal, of course, is to find a 'chapter' at which we have I,i+l(dT) = I,i(dT), a 'fixed point' at which anything dtrue in the language is fully recorded in the given chapter one needn't go further. Thinking of the various 'chapters' as *la~t9~uagcs*, each with a richer interpretation of 'dtrue', one can think of the 'fixed chapter' as a language that, finally, has a dtruth predicate for itself.

Returning to the construction at hand, we have our Strong Kleene (but classical) 'ground language' £~ that we now expand to/;t, the syntax of which includes that of £~ but also has dT(z) (and the resulting sentences formable therefrom). We want the new language to 'expand' the ground language, and we want the former to have a model that differs from the latter only in that it assigns an interpretation to dT(z). For present purposes, we let *It*, the interpretation function in £t, assign (0, 0) to dT(z), where (~, 0) is the function that assigns $\overset{1}{\sim}$ to each element of l)t. (Hence, the extension and antiextension of dT(z) in £t are both empty.) This is the formal analogue of 'chapter zero'.

---

[33] This is not to say, of course, that one can't have a Strong Kleene or paracomplete, in general language some proper fragment of which is such that A V ~A holds for all A in the fragment. (One might, e.g., stipulate that arithmetic is such that A V ~A holds.)

[34] This is usually put (more precisely) as that the domain contains the G6del-codes of all such sentences, but for present purposes I will skip over the mathematical details.

The crucial question, of course, concerns *further expansion*. How do we expand the interpretation of dT(x)? How" do we move to 'other chapters' ? How, in short, do we eventually reach a 'chapter' or language in which we have a dtruth predicate for the whole given language? This is the role of Kripke's 'jump operator'. What we want, of course, are 'increasingly informative' interpretations (T~+, T~) of dT(x), but interpretations that not only 'expand' the previous interpretations but also *preserve* what has already been interpreted. If A is dtrue according to chapter i, then we want as much preserved: that A remain dtrue according to chapter i + 1. This is the role of the 'jump operator', a role that is achievable given the so-called *monotonieity* of Strong Kleene valuation scheme n.35 The role of the jump operator is to eventually 'jump' through successive interpretations (chapters, languages) Lz(dT) and land on one that serves the role of dtruth serves as an interpretation of 'dtrue'. As above, letting I,~(dT) be a function (T~+, T~-) yielding 'both chapters i', the goal is to eventually 'jump' upon an interpretation (T~+, T~) such that (Ti+ , ~) = (T~+I, T~+I).

Focusing on the 'least such point' in the Strong Kleene setting, Kripke's construction proceeds as above. We begin at stage 0 at which dT is interpreted as (0, 0), and we define a 'jump operator' on such interpretations: 36 dT is interpreted as (~+l, ~+l) at stage i + 1 if interpreted as (T+ , T~) at the preceding stage i, where, note well, ~/;1 comprises the sentences that are true (designated) at the preceding stage (chapter, language) i, and T~+1 the false sentences (and, for simplicity, non-sentences) at i. Accordingly, we define the 'jump operator' g~ thus: 37

T +

The jump operator yields a sequence of richer and richer interpretations that 'preserve prior information' (given monotonicity), a process that can be extended into the transfinite to yield a sequence

),...

defined (via transfinite recursion) thus: as

aSMonotonicity is the crucial ingredient in Kripke's (similarly, Martin Woodruff's) general result. Let Ad and fl/V be paracomplete (partial) models for (uninterpreted) L. Let F + be the extension of F in Ad, and similarly FI+I, for Adt. (Similarly for antiextension.) Then Ad/ *extends 2M* iff the models have the same domMn, agree on interpretations of names and function signs, and 5=+ • F +, and F M • FM, for all predicates F that M and *M'* interpret. (In other words, Ad~ doesn't change 2~l's interpretation; it simply interprets whatever, if anything, Ad left uninterpreted.) MONOTONICITY PROPERTY: A semantic (valuation) scheme cr is *monotone* iff for any A that is interpreted by both models, A's being designated in 3/[implies its being designated in 2t/V whenever fl/V extends 3/[. So, the montonicity property of a scheme ensures that it 'preserves truth (falsity)' of 'prior interpretations' in the desired fashion.

36S0, our operator operates on the set of all (admissible) functions from 1:) into {1, 1,0}, where Z) is in our 'ground language', as above.

37Note that Kripke's definition applies to *any* monotone scheme c~. I relativize the operator to t~just to remind that we here focusing on the Strong Kleene case.

3sIf the reader isn't familiar with transfinite recursion, just note that it's much like ordinary ab. Base. (To+, To) (0, 0).

aS. Successor. (~-2÷l,~-a÷l) J~((7-2,7-c~)).

al. Limit. For limit stages, we collect up by unionising the prior stages:

What Kripke showed for *any* monotone scheme, and a fortiori for Strong Kleene is that the transfinite sequence reaches a stage at which the desired dtruth predicate is found, a 'fixed point' of the jump operator such that we obtain

('~2,'~0,) ('~2+1,'rc~+1) ,Jg(('~2,'rc~))

The upshot is that 'chapter a' or 'language a' is such that T+ and T~- comprise all of the dtrue (respectively, dfalse) sentences of £~, the language at a, which is to say that £~ contains its own dtruth predicate.

The *proof* of Kripke's result is left to cited work.[39] For present purposes, I move on to a few comments about the given paracomplete account.

Comments

In this section I briefly mention a few" issues concerning the philosophical application of Kripke's paracomplete account, concentrating on the unifying projects, NTP and ECP. I leave much of the discussion to cited sources, especially with respect to what I call 'interpretation issues'.

Interpretation issues

Three salient interpetation-related issues have emerged with respect to Kripke's proposal(s).[4°] I simply mention the issues here, pointing to cited works for further discussion.[4x]

---

recursive definitions except for requiring an extra clause for so-called 'limit ordinals'. Here, c~ and e are ordinals, and), a 'limit ordinal'. (One can find a discussion of transfinite recursion in most standard set theory books or metatheory textbooks. See too [McGee, 1991], [Smullyan, 1993].)

[39] Kripke's own proof is elegant, bringing in mathematically important and interesting results of recursion theory and inductive definitions. Kripke's proof is also perhaps more philosophically informative than a popular algebraic proof, especially with respect to the least fixed point (on which we've focused here). Still, if one simply wants a proof of the given result (e.g., existence of least fixed point), a straightforward algebraic proof is available, due to Visser [2004] and Fitting [1986], and discussed in a general, user-friendly fashion by Gupt~Belnap [1993].

[4°] Despite my talk of *Kripke's proposal* or the like, I should make the traditional note that Kripke himself abstained from endorsing any of the particular accounts in his [Kripke, 1975]. So noted.

[41] I should perhaps emphasize that, by my lights, the 'interpretation issues' are important and, were it not for space considerations, I would devote much more space to them.

Ikl. Which fixed point?

Ik2. Supervenience or Transparency?

Ik3. Classical or nonclassical theory?

I will (very) briefly discuss each issue in turn.

(Ikl): Take any monotonic semantic (valuation) scheme ~.[42] Kripke showed that the (suitably defined) 'jump operator' over ~-interpretations will have a fixed point that can serve as an interpretation of a dtruth predicate (for the given language). But while we've narrowly focused on one particular interpretation (the least fixed point), there are in fact many fixed points (as Kripke noted), and there's some controversy about which of the many fixed points best model the dtruth predicate. For discussion of the issue see [Gupta and Belnap, 1993].

(Ik2): This issue is related to (Ikl). Michael Kremer [1988] argues that the so-called 'supervenience' and 'transparency' ideas about dtruth are in conflict in Kripke's proposal. The former is the idea that once the non-semantic 'facts' are fixed, then so too is the interpretation of 'dtrue'. The latter idea is as indicated throughout: that 'dtrue' is entirely transparent, that it, unlike the former idea, dictates no particular interpretation other than one that affords its essential intersubstitutivity. A related issue concerning the philosophical significance of fixed point constructions, generally is discussed by Philip Kremer [2000].

(Ik3): This is a slightly more technical issue. In §4.1 §4.1 I focused entirely on r~-based languages, the Strong Kleene, least fixed point proposal. Moreover, I have focused entirely on a nonclassical reading of that proposal one for which the resulting logic is nonclassical (and, indeed, Strong Kleene). But a related, classical reading is also well-known, a reading (and resulting classical truth theory) proposed by Sol Feferman [1984]. This reading is standardly called 'KF' (for Kripke Feferman). While the proposal is interesting, it is not an account of *dtruth,* since it gives up the essential intersubstitutivity of dT(A) and A.[43] See [Reinhardt, 1986] and [Maudlin, 2004] for further discussion and related proposals.

Non-triviality project

Recall that NTP, the non-triviality project, is to explain how we can enjoy a non-trivial language that has a dtruth predicate and Liar-like sentences. In Kripke's case, the project is to show how we can enjoy a *consistent* dtruth predicate despite Liars. In general, one aims to answer NTP by constructing an artificial, formal language the *model language* that contains its own dtruth predicate, and then claims that, at least in relevant respects, 'real dtruth' is modeled by truth-in-the-model language. The point of the model language is to serve as a clear, albeit idealized, model that explains how dtruth (in our real language) can achieve consistency (or, more broadly, nontriviality). [44]

---

[42] Kripke [1975] explicitly discusses two other well-known schemes, namely, the Weak Kleene scheme (which is Bochvar's 'internal' scheme) and van Fraassen's *supervaluational* scheme. (Martin Woodruff [1975] proved the existence of the 'maximal fixed point' wrt the Weak Kleene scheme.)

[43] That it gives up on dtruth is plain from the fact that Feferman was after a consistent, classical theory, something with which dtruth is incompatible, at least given usual syntactic resources. (One conspicuous example of the departure from dtruth is in KF's commitment to instances of *A A ~Tr(A},* something that would be inconsistent if *'Tr'* were read as dtruth.)

Some standardly object that Kripke's account doesn't answer NTP, the reason being that certain notions used in the 'metalanguage' (the language in which one constructs one's 'model language') are not expressible in the object- or model language. I will return to this issue what is often dubbed *revenge* in §10. For now, I will simply sketch an example of a common charge against Kripke's (and many similar) accounts.

Concentrate on the canvassed case, namely, the Strong Kleene case. Let £~ be such a (fixed point) language constructed via the co-scheme.4s In constructing £~, we use in out" metalanguage classical set theory, and we define *truth-in-£~* (and similarly, £~-falsity). Moreover, we can prove in our metalanguage that, despite paradoxical sentences, a sentence dT(A} is /2~-true exactly if A is £~-true.

The common charge is that £~, so understood, is not an adequate account of dtruth (dfalsity) itself: it fails to illuminate how dtruth itself achieves consistency. The charge is that £~-truth achieves its consistency in virtue of/2~'s expressive poverty: L;~ cannot, on pain of inconsistency, express certain notions *that our 'real language' can express.* Example: Suppose that £~ contains a formula *A(x)* that defines {B : B is not £~-true}. And now, where ~ says A((A)), [46] we can immediately prove in the metalanguage that A is £~-true iff A((A}) is /2~-true iff A is not £~-true. Because and only because we have it *in our classical metalanguage* that A is £~-true or not, we thereby have a contradiction: that A is both £~-true and not. But since we have it that Z2~ is consistent (given consistency of classical set theory in which Z2~ is constructed), we conclude that £~ cannot express 'is not £~-true'.

The common charge, then, amounts to this: that the Kripkean model fails to be 'enough like our real language' to explain at least one of the target phenomena, namely, dtruth's consistency. Our metalanguage is part of our 'real language', and we can define {B : B is not /2~-true} in our metalanguage. As the Kripkean language cannot similarly define {B : B is not £~-true}, the Kripkean model thereby fails to illuminate dtruth's consistency. In short, the proposed model fails to adequately answer NTP.

Despite the popularity of this sort of charge, I think that it is confused. I will return to this in §10, but the basic confusion arises from conflating the model-relative notions (e.g., truth-in-the-model language), which are defined in a classical metalanguage, and the target notions in the 'real language'. Since classical logic is an extension of Strong Kleene semantically put, any classical interpretation is a Strong Kleene interpretation there's no obvious reason why one can't ~stand squarely within a classical fragment' of one's 'real language' and define model-relative notions that serve as heuristic tools (serve, e.g., to answer NTP).[47] If that is correct, then one can't expect 'true-in-the-model language' to be exactly like dtruth, since on the paracomplete proposal LEM fails for the latter but holds for the former. Indeed, the difference between the two notions is conspicuous: dtruth cannot be defined in a classical language (or fragment thereof), but 'true-in-the-model language' can be so defined (e.g., in Kripke's and others' proposals). So, even on the surface, there is reason to think that the common charge (sketched above) turns on a confusion.[48]

---

[44] Moreover, and importantly, such artificial 'model languages' aim to serve as *models* in a more technical sense affording a *consistency proof.*

[45] The point applies to any of the given languages.

[46] This is the familiar sense of 'says' introduced in [Burge, 1979].

Put another way, the problem with such common charges is that they're either confused or unwarranted. The paracomplete theorist proposes that LEM fails for negation (and, presumably, any 'negation-like' devices) in our 'real language', the language for which NTP arises. In the Strong Kleene framework, the paracomplete theorist claims that Strong Kleene is our 'real logic', the logic governing our real language. But that's compatible with the real language having an entirely classical proper fragment the logic of which restricted to that fragment is classical. (Recall, as above, that classical logic is a proper extension of Strong Kleene.) For convenience (or, perhaps, other reasons), the paracomplete theorist, towards answering NTP, 'stands squarely in the classical fragment' of her real language to construct an artificial language a model language intended to model the real language in relevant respects. The construction consists of defining various model-relative notions such as *true-in-the-model language* and, note well, *its classical complement* (viz., 'not true-in-the-model language') for which, of course, LEM holds. Now, the common charge, as sketched above, is *confused* if it conflates the model-relative notion of 'not true-in-the-model language' with, for example, 'not dtrue'. But it should also be plain that the charge is *unwar'ranted* if it merely assumes that there's some negation-like device *No T* in the real language for which, for example, every sentence (in the real language) is either dtrue or NoT. The paracomplete theorist maintains that, perhaps in addition to other phenomena, the paradoxes teach us that there's no such device. The common charge, in turn, must come equipped with an argument that there is such a device, if it is wishes to show the inadequacy of the paracomplete account with respect to NTP. (Pointing out that we use 'not true-in-the-model language' in our real language is irrelevant, since that, as above, is merely model-relative.) So, such common charges are either unwarranted or confused.[49]

---

[47] In terms of the logics, one can think of a logic being an extension of another iff the former 'contains more validities'. More precisely, letting VL and *VL,* comprise the valid sentences and 'rules' (argument-validities) of two different logics L and L~, we say that L~extends L exactly if VL, • VL, and *properly extends* iff *VL,* C VL. As in §4.1, it is easy to see that classical logic (properly) extends Strong Kleene. (What mightn't be as obvious is that all classically valid rules that don't turn on classically valid sentences are valid in Strong Kleene.)

[48] Hartry Field [2003b], correctly in my opinion, conjectures that some of the noted confusion, at least targeted at the Kripkean Strong Kleene proposal(s), may well arise due to a confusion between the nonclassical reading, on which I've focused here, and the KF-reading noted above. On the latter reading, the proposal has it that some sentences are *neither true nor false.* Such a notion makes little sense if 'true' and 'false' are understood as *dtruth* and *dfalse.* (This, of course, is precisely not how 'true' or 'false' are understood in KF!) On the nonclassical reading, in which we have dtruth, it makes little sense to talk in terms of 'neither dtrue or dfalse'. I will return to the issue of 'neither dtrue or dfalse' in the next (sub- sub-) section, and also in §4.2.

Despite the confusion or lack of warrant in such common charges, there is none the less something correct in the objection, something about expressive poverty. But the issue, I think, has less to do with NTP than with ECP, to which I now (briefly) turn.

E:dtaustive characterization project

EcP is the project of explaining how, if at all, we can achieve 'exhaustive characterization' in a language with its own dtruth predicate (in the language) and Liar-like sentences. While 'exhaustive characterization' remains imprecise, the intuitive import is clear: a language in which we have various 'semantically significant predicates' that may be used to *e:dtaustively* and correctly (semantically) categorize all sentences of the language.

One way of thinking about 'exhaustive characterization', as here intended, is as follows. Suppose that our semantic-free fragment/2o is exhaustively characterized *classically,* in which case we have it that, where cEc is the 'classical exhaustive characterization',

cEc. Every sentence of/20 is either dtrue or dfalse.

Now, what Kripke showed   and paracomplete theorists, in general, advance   is that we can retain our dtruth device (the predicate 'dtrue') if our language goes beyond the classical   opens up the 'semantically significant options' for sentences. The idea is that, while cEc may suffice for Z:0, a genuinely *e:chaustive* characterization requires another category:

Ec. Every sentence (in the language) is either dtrue, dfalse, or *Other.*

The question is: how"shall *Other* be understood?

To begin, one way that 'Other' should *not* be understood is as implying *not dtrue* or, hence, *neither dtrue nor dfalse,* at least in paracomplete accounts in which normal De Morgan principles (e.g., distribution-like principles) hold for the extensional connectives.[5°] Suppose, for example, that 'other' in Ec is cashed out such that Other(A} implies ~dT(A}, and hence   given intersubstitutivity   implies ~A. One reason for introducing 'Other' into the language is to *correctly characte~ize* Liars   e.g., the ticked sentence in §3. Let A be sentence that says of itself (only) that it's not dtrue, that is, a sentence equivalent to ~dT(A) and, hence, ~A. If 'Other' is to play the role for which it was introduced namely, to correctly characterize, perhaps among other things, Liars then one would want Other(A} to be dtrue. But if, as supposed, Other(A) implies ~dT(A) and, hence, implies ~A, then inconsistency abounds: from the fact that A is *other,* it follows that A is *not dtrue,* in which case, since that is precisely what A says, A is also dtrue, by the essential intersubstitutivity of dtruth.

---

[49] Hartry Field [2003b; 2005b] has made a lot of use   good use, in my opinion   of this basic point. But I trust that the main point is fairly clear and uncontroversial. (Where controversy might arise is with EcP, to which I turn below.)

[5°] In my discussion of paracomplete (and, for that matter, paraconsistent) accounts, I focus entirely on 'normal accounts', accounts for which conjunction, disjunction, and negation interact in standard ways   at least for the 'classical values'. While I think that *non-truth-functional* approaches are worth exploring, e.g. [Beall, 2005a], I omit those here due to space considerations.

In general, then, it is not difficult to see that 'Other' in EC cannot be consistently understood as implying *not dtrue) 1* The point applies in particular to the Strong Kleene proposal of Kripke: it makes no sense to say of Liars that they are neither dtrue nor dfalse) 2

In the end, while the account provides a (paracomplete) response to NTP, Kripke's Strong Kleene proposal affords no answer to ECP.5a As a result, there's no 'significantly semantic predicate' that is introduced in the given language (as opposed to model-relative metalanguage-terms) for purposes of correctly characterizing Liars, at least on Kripke's paracomplete account. And this is thought by some, perhaps many, to be an inadequacy of the proposal.

I will return to the issue of ECP below. For now, I turn to one more issue concerning Kripke's Strong Kleene proposal, the issue of a suitable conditional.

Dtruth and dT-biconditionals

What are standardly called T-biconditionals, biconditionals of the form T(A} ~ A, have long been thought to be an essential feature of truth, something at least essential to 'the' so-called naive theory of truth. But what about *dtruth?* What is essential to dtruth is its transparency, its intersubstitutivity. Whether all instances of dT(A) ~ A hold depends entirely on the sort of conditional one has, whether, for example, one's conditional ~ is such that A ~ A is valid. If A ~ A is valid in the language, then (obviously) the dT-biconditionals will thereby hold (assuming, as I will, that ~ is defined as usual).

Whether an acceptable theory of dtruth must be one that validates all dT-biconditionals is an important question that I'll none the less leave open. For

---

51Recall that we are focusing on *dtruth,* which requires the intersubstitutivity of *dT(A}* and A. Deviating from dtruth affords more options, and in many ways is what (or, more accurately, what may have) motivated approaches like [Feferman, 1984], [Maudlin, 2004], and so on.

52Recall the difference between the *model language* and the *real language,* and in particular the difference between *model-relative* (ultimately, merely instrumental) notions and the 'real notions' intended to be modeled. It makes perfect sense, *standing squarely and only in the classical fragment (metalanguage),* to say of sentences in the model-language that they're neither true-in-the-model language nor false-in-the-model language. But such model-relative notions aren't at issue here.

5aIn fact, one might take some of Kripke's remarks to suggest that, by his lights, there's no escaping a 'Tarskian' or 'hierarchical' approach to ECP. (See the famous passage about Tarski's ghost [Kripke, 1975].) But I will not dwell on the exegetical issue here. [Note that one might take the 'classical reading' of Kripke, formulated by Feferman's KF, as having an answer to ECP, and in many respects that's correct. (One can truly say, e.g., that Liars are not true.) But, again, KF gives up dtruth (and, so, NTP wrt dtruth).]

present purposes, I will assume without argument that, other things being equal, an account of dtruth that validates all dT-biconditionals is prima facie more attractive than an account that fails to do so. Accordingly, an account of dtruth for which we have the validity of A --~ A is prima facie more attractive than an account for which A ~ A isn't valid (at least other things being equal). 54

That A ~ A is not valid in Kripke's Strong Kleene proposal is clear, since 4, in Strong Kleene, is simply the material conditional, which is defined ~A VB, and here abbreviated as A D B. But, then, A D A is valid only if ~A VA is valid, that is, only if LEM is valid. But LEM isn't valid in Strong Kleene, as noted in §4.1. Indeed, the heart of paracomplete proposals is the rejection of LEM.

While Kripke's proposal seems to show how we can have a nontrivial (indeed, consistent) dtruth predicate despite the existence of Liar-like sentences, it fails to show how we can achieve as much in a language for which the dT-biconditionals hold.

One might think it an easy fix to add a conditional. After all, Lukasiewicz's (pronounced ~woo-kush-YE-vitch') 3-valued language differs from the Strong Kleene language only in that it adds a conditional for which A ~ A is valid. (One can retain the hook, of course, so as to have two conditionals, --+ and D. This is no surprise given that, as noted above, A D /3 simply is ~A V/3.) Lukasiewicz's conditional is defined thus: 55

43	1	'r~	0
i	i	~	0
~	1	1	'r~
0	1	1	1

As one can see, A 43 A is always designated in the Lukasiewcz semantics, and hence, given the essential intersubstitutivity of dtruth, dT(A) +-+3 A is always designated (with +-+3 defined as usual via conjunction).

The *trouble,* however, is that this proposal will not work in the sort of paracomplete, fixed-point languages at issue. One way to see this is to consider a version of *Curry's paradox. 56* Without getting into the technical details, a simple way to see the problem is via an informal Curry-like situation. Assume a Strong Kleene (fixed point) language augmented with the Lukasiewicz conditional above. Let A be a sentence that says dT(A} ~+a ±, where ± is some false sentence in the semantic-free fragment, say, ~1 = 0'. Let C be a sentence that says dT(C) +-+3 dT(A). A paracomplete theorist will want to say that A is to receive value 1 (or modeled

---

[54] One natural route towards an *argument* is given by Feferman [1984], who argues that without the validity of A --+ A 'ordinary reasoning' is crippled.

[55] Here, n = 5"

[56] See [Curry, 1942] and, for relevant discussion, [Meyer *et al.*, 1979]. Note: Some authors [Barwise and Etchemendy, 1987] call Curry's paradox *LSb's paradox,* mostly due to the similarity between LSb's Theorem (or the proof thereof) and Curry's paradox. (The situation is somewhat similar to GSdel's Incompleteness proof and the Liar paradox.) But Curry certainly discovered the paradox much earlier and, indeed, it's likely that Medieval logicians discovered it prior to Curry.

as such). Suppose that C receives the value ½. Then the values of C and A are the same, in which case C gets the value 1. Contradiction. Similarly, a contradiction arises if C receives 0 or, obviously, 1. Hence, there's no obvious way to add Lukasiewicz's conditional to the Strong Kleene (fixed point) language.[57]

Curry's paradox, in general, imposes constraints on conditionals. A *Curry sentence* is one that says of itself (only) that *if* it is dtrue, then everything is dtrue.[5s] So, a Curry conditional is a sentence C equivalent to C ~ ±, where ± is 'explosive', a sentence implying triviality (e.g., 'Everything is dtrue'). Let us assume that a 'genuine conditional' is one that *detaches* in the following form (sometimes called 'rule-form'), where IF- is our consequence relation (either semantically or proof-theoretically defined).

A,A ~ B IP B

And assume that we want a genuine conditional such that A ~ A is valid (and, hence, given dtruth, the dT-biconditionals are valid). Then to avoid Curry paradox, one must either reject *reductio* or *contraction* in the following (among other) forms.[59]

R. A~AIP~A

C1. A A (A ~ B) ---+/3

C2. (A ~ (A ~ B)) ~ (A ~ B)

C3. *A--+ (A--+ B) I[- A--+ B*

Consider, for example, (C1). Assume that we have a conditional --~ such that A --~ A is valid (and, hence, the dT-biconditionals), and also have both dT and Curry sentences in the language. Let C be such a sentence, equivalent to dT(C) ~ ±, in which case, via the dT-biconditionals and intersubstitutivity, we have

. c (c ±)

---

[57] This is slightly misleading, as put, but without getting into a much more technical discussion, I will leave the general problem there.

[58] The consequent of a Curry conditional may be any sentence that implies triviality or near-enough triviality (e.g., '1 = 0'), where a language (or truth predicate) is trivial iff everything is true (in the language).

[59] I should also note that in *paracomplete* languages in which A ~ A is valid and ~ is 'genuine' (and one has a dtruth predicate), one must also reject Importation and 'Introduction' in the following forms.

Imp. *A~(B~C) IF-AAB~C*

Int. From A I~/3 infer I~A –+/3

The import of rejecting (Int), so put, is that 'deduction theorems' won't generally hold. (I briefly discuss a related issue in §6.) It's worth noting that the otherwise apparent obviousness of (Imp) rests largely on assumptions of contraction-principles that, for Curry reasons, must be rejected. A nice discussion of these issues can be found in 'relevant logic literature' [Anderson and Belnap, 1975; Anderson *et aL,* 1992; Routley *et al.,* 1982] and, of direct relevance here, Field's [2002; 2003a]. See too [Restall, 2000].

But, then, from (C1) we get

c A (c ±) ±

which, via substitution, yields

CAC~±

which, assuming normal conjunction, is equivalent to

** C~±

Given that ~ is a 'genuine conditional' and, so, detaches, we quickly get C itself via (*), and now" detaching again via (**) yields Z. Triviality.

One might think that, while his 3-valued conditional won't work, the full continuum-valued language £o~ of Lukasiewicz, which, except for more values, retains the (K1) (K3) clauses for standard connectives (see §4.1), might do the trick. In particular, 4oo, the conditional in £o~, satisfies none of the given contraction principles nor (R).6° Unfortunately, the proposal won't work, as Greg Restall [1992] and, more generally, [Hajek *et al.*, 2000] showed: the resulting theory will be c~-ineonsistent. ~1

4.2 *Field*

The upshot of §4.1 is that while Kripke's paracomplete proposal shows that, despite having Liars in the language, we can have a nontrivial (indeed, consistent) dtruth predicate (in the language, for the language), it none the less exhibits two apparent inadequacies.

I1. The proposal fails to answer NTP for a language in which all dT-biconditionals hold.

I2. The proposal fails to answer ECP in any fashion.

As in §4.1, Curry's paradox puts constraints on adding a genuine conditional for which A ~ A is valid (and, hence, for which the d-r-biconditionals all hold). Answering ECP, in turn, likewise requires care, since 'revenge Liars' are ever-ready to emerge.

Hartry Field, in a series of papers (see throughout), advances the Kripkean paracomplete proposal by attempting to overcome (I1) and (I2), the two notable inadequacies of Kripke's own proposal. Field maintains the basic paracomplete line that paradoxes (perhaps among other phenomena) teach us that LEM is to be

---

6°More promising yet, perhaps, is that naYve 'property theory' (sometimes called 'na'l've set theory') is consistent in £~. See [White, 1979].

61I will not get into the details of formulating, e.g., Peano Arithmetic in £~ and showing the resulting w-inconsistency. For details, I recommend beginning with [Restall, 1992] and, in turn, the more generalized [Hajek *et al.*, 2000].

rejected, that some instances of $A \vee {\sim}A$ are to be rejected.[62] Field's contribution is an answer to NTP for a language in which all dT-biconditionals hold and, in turn, an answer to ECP.[63]

Field's main contributions, as above, are his conditional and his approach towards characterization (classifying Liars, etc.), with the latter nicely falling out of the former.[64] Given the aims of this essay, I will only provide a sketch of (the basic idea of) the conditional and, in turn, Field's approach towards characterization.[65] But before sketching the conditional, I will first (briefly) discuss the background philosophical picture, and then turn to (a sketch of) the formal model.

Philosophical picture: stronger truth

An inadequacy of Kripke's proposal is that we're left with nothing to truly say about Liar-like sentences, at least not *in our language* (the language that enjoys its own dtruth predicate).[66] Intuitively, the paracomplete theorist thinks that, for purposes of truly characterizing or 'classifying' Liar-like sentences, we need to acknowledge an additional 'semantically significant' category beyond dtruth (and dfalsity). But how" shall this 'other' category be understood?

As in §4.1, there's no clear sense in saying that Liars are *neither dtrue nor dfalse*. Still, one might think that there's *some* sense in which Liars are 'not true' or 'not false'. This thought motivates Field's proposal.[67]

---

[62] This is not to say that there aren't significant (proper) fragments of the language for which LEM holds. For example, one might maintain that LEM holds over the arithmetical fragment of our language, or physics, or so on. (Field [2003b] explicitly agrees with this, suggesting that LEM may hold for mathematics or the like, in general.) Of course, as mentioned in §4, since we're (here) dealing with dtruth, *rejection of* A is not going to be *acceptance of* ${\sim}A$. Field naturally suggests weakening the classical 'exhaustion' rule for probability (or degrees of belief), namely, $Pr(A) + Pr({\sim}A) = 1$, to one that respects a paracomplete approach: namely, $Pr(A) + Pr({\sim}A) \leq 1$. There are subtle issues here, but for space I leave them aside.

[63] The respect in which Field answers the intended import of (the admittedly imprecise) ecP is something I'll briefly discuss below. What I should emphasise is that, in my opinion, Field's work is a remarkable advance in the area of *dtruth* and paradox. (And, of course, Field has also contributed a great deal to the philosophical issues concerning *dtruth*. See]Field, 2001].)

[64] Actually, Field should also be given credit for explicitly noting that paracomplete accounts of rejection (and acceptance) call for an adjustment of classical probability theory. While this point is often assumed, Field took the time to explicitly propose a revision, one that, by my lights, seems to be entirely right (for purposes of a paracomplete account). See]Field, 2003b] for a nice discussion.

[65] Field continues to revise the proposal, coming up with further refinements and improvements. My aim is not to give the very latest, or even the full details of any particular version. Rather, my aim is merely to sketch the basic idea.

[66] Kripke precisely defines notions of *ungrourzdedness* and *paradoxical* (and more). One might think that we can truly characterize Liar-like sentences as *paradoxical,* so defined (per Kripke). But so defined, one can't consistently add 'is paradoxical' to the language, at least if it is to behave as Kripke seems to suggest e.g., as a predicate the extension and anti-extension of which are exhaustive and exclusive. It is precisely that assumption that paracomplete theorists ought to reject! And Field recognizes as much.

[67] The thought likewise seems to motivate a related (and well-known) earlier proposal by McGee [1991]. McGee's work precedes that of Field's, but the two are related.

The paracomplete theorist rejects that Liars are dtrue or not dtrue (or, equivalently, dfalse or not). But perhaps one can recognize a stronger notion of 'truth' according to which Liars and their negations are *not true.* Let sT be our 'stronger truth predicate' (stronger than dT). Being a stronger notion than dtruth, one might have it that while both

st1. IF sT(A) --~ A for all A (and for some suitable conditional!)

st2. sT(A} IF A for all A

hold, the converse of either (st1) or (st2) fails.[6s] This failure need not get in the way of expressing generalizations or the like, since *dtruth* remains as before full intersubstitutivity holds. Dtruth remains our expressive device. The new device, sT, is brought in to do a job that dtruth was never intended to do: namely, *fail to be transparent!* In particular, we want a device that allows us to truly (dtruly) 'classify' Liars, sentences like the ticked sentence in §3. For reasons above, we can't classify the ticked sentence oi" its negation as dtrue (or, hence, dfalse); however, with our stronger notion, we may be able to classify the sentence and its negation as *not strongly true.* That is the idea.[69]

Care, of course, must be taken. Obviously, on pain of 'revenge paradox', sT, however it is spelled out, must be such as to resist excluded middle, resist having sT(A} V ~sT(A}. (Just consider a sentence A equivalent to ~sT(A}.) And this is precisely in keeping with the paracomplete theorist's rejection of LEM: that there is no 'truth-like' or 'negation-like' device t (be it a predicate or operator) in the language that is 'exclusive' in the sense of satisfying both of the following.

el. IF tA V A for all A.

e2. ~A, A IF B for all A and B.

The heart of paracomplete accounts is that at least (el) fails for our language.[7°]

But now an apparent tension arises. The reason that we want a stronger notion of truth is that we want to be able to dtruly classify Liar-like sentences, 'characterize' them as being *in some sense* 'not true'. So, we bring in sT, and we want to say for any Liar-like sentence A in the sT-free fragment that A is not strongly true; we want to assert ~sT(A). This much is not difficult, provided we're restricting ourselves to As in the sT-free fragment. The *trouble,* of course, is that we want to talk about *arty* sentence in the *full* language, including any sT-ful sentences sentences that use 'sT'. But, then, we want to be able to say of sT-ful Liars that they are not strongly true. And that's the problem. Inevitably,[71] there will be

---

[6S] Typically, e.g. [McGee, 1991], the converse of (st1) fails. This is likewise the case for Field's proposal.

[69] Such 'strong truth' is often understood in terms of *determinately dtrue,* the idea being that by rejecting each of A and ~A, one is rejecting that either sentence is determinate.

[7°] Note that we're not talking about *model-relative* notions that may be defined in an entirely classical, *proper* fragment of one's full, paracomplete language. We're talking about 'non-model-relative' or, as Field (in conversation) says, 'absolute' notions.

[71] Formally, the force of 'inevitably' is ensured by GSdel's diagonal lemma.

sT-ful sentences A such that, were either sT(A) or ~sT(A) to hold, inconsistency would follow. Avoiding such inconsistency requires, as above, the rejection of (the equivalent of) excluded middle for sT. But, now', there would seem to be sT-ful sentences that fail to be 'correctly classified' by sT.

The apparent tension can be seen as follows.[72] We *begin* with a transparent expressive device dT, which, when introduced into the language (to play its transparency role), gives rise to dT-ful Liar-like sentences. The paracomplete theorist maintains that such sentences are not problematic if we reject LEM, and in particular (at least) the Liar-instances of LEM. So, we can keep our (consistent) expressive device dT despite its inevitable Liars. One problem solved. But, *next,* we want to be able to dtruly 'classify' or 'characterize' the given Liars. Towards that end, we introduce a stronger notion of truth, sT.[73] And now we can use sT to dtruly classify all dT-ful Liars. Another problem solved. But *now again* to avoid inconsistency, we must likewise reject LEM for sT, and in particular reject sT-ful Liar-instances of LEM. We now seem to require *yet another* 'even stronger truth' to dtruly classify the given sT-ful Liar-sentences (ones remaining 'unclassified' on pain of inconsistency). And so on, ad infinitum.

And now the tension is clear. Suppose that we have some *unified predicate* (say, 'Other') that characterizes all Liar-like sentences, so that we have the following.

EO. IF dT(A} V dF(A) V Other(A)

Then we seem to be stuck in inconsistency, at least assuming normal behavior for the (extensional) connectives.[74] After all, consider a sentence A that says ~A V Other(A). By EC, A is either dtrue, its negation dtrue, or it is Other. If dtrue, then A is dfalse or Other. If dfalse or Other, then A is dtrue. If, then, the given semantically significant 'categories' are exhaustive and exclusive, then inconsistency arises.[75]

What the picture of infinitely many 'stronger truth predicates' requires is the absence of any unified (semantically significant) predicate in terms of which all Liar-like sentences are to be classified.[76] The given picture must be accompanied by a rejection of anything yielding something along the lines of EC.

Perhaps the rejection of EC the rejection of a unified (semantically significant) predicate in terms of which all Liar-like sentences are classified is not unnatural. After all, the heart of standard paracomplete accounts is a rejection of LEM *all the way through.* In broadest (though, admittedly, somewhat vague) terms, the rejection of LEM might be seen as a basic rejection of an 'exhaustive characterization' in terms of unified (semantically significant) predicates.

---

[72] I will assume a Strong Kleene approach to the basic, extensional connectives. This is the case with all paracomplete accounts that I discuss Kripke, Field, and the 'paranormal' suggestion.

[73] Here, sT can be either a predicate or operator. The difference is minimal when our basic truth predicate is dtruth.

[74] But see §4.3.

[75] Again, unfortunately I remain vague about what counts as 'semantically significant', and I believe that the project of clarifying this notion is an important (and pressing) one. For now, one can think of 'dtrue' and 'dfalse' among the lot, and then any predicates in terms of which Liars are to be 'classified'.

[76] Note that this immediately requires rejecting the coherence of 'quantifying over the hierarchy' of such predicates in the otherwise intuitive sense of, e.g., *true in some sense or other of %trongly true'* or the like.

The basic tension, in the end, is one arising from a common aim: the aim of dtruly classifying (within one's language) all Liar-like sentences in one's given language.[77] Intuitively, the aim is (presumably) to achieve such exhaustive characterization using 'unified' semantically significant predicates, something that yields the likes of EC. But another route is available: classifying any given Liar via infinitely many 'stronger and stronger' truth predicates, none of which afford a 'unified predicate' that, as it were, serves as a 'unionizer' of all such predicates. While Field does not achieve the former, he none the less provides a powerful approach along the latter lines.

Formal model: conditional and determinacy

As above, Field aims to retain a consistent dtruth predicate but, going beyond Kripke, also have all dT-biconditionals *and* a way of 'classifying' any given Liar-like sentence in the language. Field shows how to add a suitable conditional to the Kripkean Strong Kleene framework, and then defines 'stronger truth' or, as Field says, 'determinate truth', in terms of the given conditional. Given that, as in §4.2, any such 'stronger truth' predicate (or operator) must resist LEM, Field's aim of 'characterizing' Liars requires infinitely many such 'stronger truth' devices. One notable feature of Field's framework is that the requisite infinite stock of (stronger and stronger) truth devices falls out of Field's constructed conditional.

This section presents only a sketch of Field's basic idea for introducing a suitable conditional into a paracomplete and otherwise merely Strong Kleene language with a (consistent) dtruth predicate. I first present an initial sketch of (the basic idea of) how to extend Kripke's initial construction with a (non-truth-functional) suitable conditional what Field calls a *restricted semantics.*[7s] In turn, I sketch a more general setting (what Field calls 'General semantics') for the conditional.[79]

Restricted semantics

Let/(3 be the Strong Kleene logic. Field's aim is to give an extension of/(3 that, in addition to containing a consistent dtruth predicate, validates all dT-biconditionals in such a way that Curry paradox is avoided. The basic proposal is a novel combination of ideas from Kripke [1975] and *revision theorists* [Herzberger, 1982; Gupta and Belnap, 1993].

---

[77] Of course, we're talking about languages that have their own, nontrivial (indeed, in this section, consistent) dtruth predicate.

[7S] Strictly speaking, Field [2003a] does show how one can expand the values of the language in such a way as to view the given conditional as 'truth-functional', but I will ignore that here.

[79] I should note (again) that Field has modified his account in various ways, with the latest [2005c] being a variation of Stephen Yablo's work [2003]. I would've liked to discuss Yablo's work, in general, as well as Field's work in more detail; however, space limitations prohibit doing so. (Field's work is cited throughout. For Yablo's work, see especially [Yablo, 1993a; Yablo, 1993b].)

We start with a (first-order) syntax supplemented with dT and a primitive two-place connective ~ (which I'll call *the conditional).* (Any sentence the main connective of which is the conditional will be called a *conditional.)* With respect to the conditional-free fragment, the language is interpreted exactly along the lines of Kripke (as in §4.1). The challenge is to interpret all sentences, including all conditionals, in such a way as to retain a (consistent) dtruth predicate and, so, achieve full intersubstitutivity even with respect to (and 'inside of') conditionals and validate all dT-biconditionals.

Field's proposal is to interpret the language via a transfinite sequence of Kripkean (Strong Kleene) fixed points P~ (for 'point c~', with c~ an ordinal), s° where each such fixed point is 'built from' an initial *starting valuation ,9~* (for 'start c~'), which assigns elements of {1, s,0} to *all and only conditionals.* Beginning with such 'start points', Kripke's construction (see §4.1) yields a value for every sentence in the language in such a way that dtruth (transparency, intersubstitutivity) is preserved. With respect to such start points S~, and in particular how any given $~ is determined on the basis of 'prior' Kripkean 79~, Field proposes the following recipe, st

Fb. Base (Zero). *$o(A ~ B)* s for all A and B.

Fs. Successor. At successor points (or stages), we look back at the prior Kripkean fixed point:

1 if 7)a(A) _<7)a(B)
Sa+l(A ~ B) 0 otherwise.

F1. Limit. At limit points (stages), we look backwards at all prior Kripkean fixed points:

1 if 7)'~(A) _<7)'Y(B) for some ~ < A
and any 7 such that ~ < -y < A
$x(A --+ B) = { 0 if 7)~(A) > 7)~(B) for some ~ < A
and any 7 such that ~ < -y < A
§ otherwise.

So goes the construction of 'start points' from the 'prior' Kripkean fixed points. As above, the latter -- the Kripkean fixed points -- are the points that yield the 'ultimate values' in terms of which all sentences eventually stabilize into a language with both dtruth and all dT-biconditionals. The various P~, as above, are determined entirely by the $~ (which give values to the conditionals) and

---

8°Note that these are fixed points of Kripke's 'jump operator', which will henceforth be left as implicit.

81For ease, I will write '8a(A)' to abbreviate *the value of A in the start point 8a.*

the Strong Kleene (minimal) fixed point construction the various clauses for compounds, (Zl) (K3). (See §4.1.)

Brief reflection on (Fb) (F1) indicates that values assigned to (at least typically paradoxical) sentences at the various 8~ and, in turn, 79~, fluctuate quite a bit; such sentences exhibit jumpy instability. By way of settling on 'ultimate values', by way of bringing about order to such apparent chaos, Field takes a leaf from revision theory (see §8). In particular, Field defines the *ultimate value of A*, say **IAI,** thus:

IAI r limJ)~(A) if the limit exists
1
5 otherwise.

In other words, if for some point a such that for any/3 > a we have it that P~(A) is 1, then IAI 1. Similarly for 0. Let n c {1,0}. The idea is simply that if A is eventually forevermore assigned n, then n is A's ultimate value. But, of course, there may be no such point beyond which the value of A 'stabilizes' at either 1 or 0 , in which case **IAI** = 1 5". More precisely, **IAI** = 1 5 if either A is never eventually forevermore assigned anything or there's some point c~ such that for any/3 _> c~ we have it that P~(A) 1 (i.e., A is eventually forevermore assigned ½).

Field [2003a] proves that such 'ultimate values' obey the K3-rules for extensional connectives, that is, for all connectives except the non-extensional --% the conditional, s') Moreover, he shows that the construction validates all dT-biconditionals, s3

I will discuss the virtues of Field's proposed conditional, and its role with respect to 'stronger truth', in §4.2. For now, I turn to a more general, and perhaps philosophically more 'intuitive', account of the conditional.

General semantics

Because other approaches to dtruth (e.g., paraconsistent) often invoke a 'possible worlds' framework for purposes of modeling a suitable conditional, I will briefly sketch Field's 'general semantics' for his conditional, a semantics that is related to possible worlds approaches, though the intended philosophical *interpretation,* as Field [2003b] remarks, is better thought of as 'possible assignments relative to actual conditions or constraints'. In this (sub-) section I will simply sketch the construction, leaving comments to §4.2. 84

The aim, once again, is to give an extension of K3, but in this case we work with a 'modalized' /(3 (first-order) language, where, as above (throughout), we can stipulate *that* the 'semantic-free' fragment is entirely classical, and the dtruth

S2More technically, Field's 'Fundamental Theorem' shows that there are ordinals "7 such that for any (non-zero) /3 the value of any sentence A at "~•/3 is just IAI.

SaField also shows that the dT-biconditionals are validated in a stronger sense that he dubs 'conservative', the idea being (roughly) that the resulting theory is consistent *with any arithmetically standard starting model.* For discussion, see [Field, 2003@

S4A few caveats: first, the 'general semantics' was motivated largely towards a unified solution to both semantical and soritical (vagueness-) paradoxes. I will not discuss the latter here. Second, I will not be giving all of the constraints that Field proposes for purposes of achieving various desirable features (of the conditional); I will simply sketch the basic idea.

predicate is achieved along Kripkean lines, as above, s5 The difference, now, is that we expand our interpretations with a (nonempty) set gf of 'worlds' and, in turn, assign values to sentences relative to such worlds, where the values remain either 1, 0, or ½. This much is standard. The task is to tweak the interpretations in such a way as to give the target conditional its desired features.

Field's proposal is a novel variation on so-called 'neighborhood semantics'.S° We let W be an infinite set of worlds at which sentences are assigned an element of {1, ½,0}, letting @ be a (unique) distinguished element of W, the 'actual world'. In turn, we impose a 'similarity relation' on gf in such a way that each w c W comes equipped with a set of 'sufficiently similar worlds' (a so-called 'neighborhood' of w), worlds that satisfy some condition of similarity with respect to w. Specifically, Field proposes that each w E W be assigned a (possibly empty) *directed* family 5~ that comprises nonempty elements of p(W), nonempty subsets of W. The *directedness* of 5w, which amounts to

(w, ~ *w)(vx, y ~ .,%)(~z ~ .,%)2 c_why*

allows for 'incomparability', that is, that the relation of similarity needn't be linear, s7

With an eye towards semantical paradox, a few" other tweaks are required, ss Define, for any w c W:

Normality. *w* is *normal* iff w c X for all X c bt-~.

Abnormality. *w* is *abnormal* iff it is not normal.

Loneliness. *w* is *lonely* iff {w} E 5c~.

Happiness. *w* is *happy* iff it is not lonely.

Field stipulates that @ be both normal and happy on any interpretation, but otherwise worlds may be abnormal and lonely, s9 Accordingly, every interpretation is such that, per normality, ~ E X for all 2(E 5c~ and, per happiness, {(~} ~ brce. Hence, ~ is 'sufficiently similar' to itself on all interpretations, and (~ is also 'sufficiently sinfilar' to some w ¢ @ on all interpretations.

As above, sentences are now assigned a value at each world. With respect to the conditional-free fragment, the valuations simply follow" the Strong Kleene rules;

SSAs noted above, the framework is motivated not only by semantical paradox but also soritical. As a result, if we were concentrating on vagueness, we wouldn't stipulate an entirely classical 'semantic-free' language.

S6The generalization of standard Kripke models is also sometimes called *Montague semantics*, or *Montague-Scott* semantics. For general references, see [Chellas, 1980].

SrField [2003b] notes that one could, without radical deviation from the basic proposal, simply impose a linear ordering via C.

SSActually, achieving all of Field's desired features requires other constraints on interpretations, but for present purposes I skip over them.

S9That Œ is to be happy is not just warmheartedness on Field's part. If Q were allowed to be lonely, various (contraction-related) validities would emerge that would engender inconsistency in the resulting dtruth theory.

(K1) (K3), in effect, are modified only with respect to being relativized to worlds, even though reference to worlds in the clauses for extensional connectives makes no essential difference. (In other words, where A is conditional-free, the value of A at w depends only on the values of A's constituent parts at w. One needn't 'look at other worlds' to figure out the value of purely extensional sentences.) The worlds come into play with conditionals.9°

1 if [A[~,_< [B[~, for someXE~c and anyw'EX
[A --0 B[~ = { 0 if [Alw, > [Biw, for some X E ~,w and any w' E X
1 otherwise.

With valuation-conditions in hand, the (semantic) consequence relation IF-may be defined. Towards that end, let us say that, relative to an interpretation, a sentence A is *actually verified* iff IAIc~ 1 (in the given interpretation). Similarly, a set E of sentences is *actually verified* iff t3 is actually verified, for each B E E. Then the (semantic) *validity* relation It- is defined thus:

E IP A iff any interpretation that actually verifies E actually verifies A

with *valid sentences* being consequences of 0. Given the existence of abnormal and lonely worlds (or the existence of interpretations containing as much), other notions of validity may be introduced, but for present purposes I will focus just on the given notion.91

So goes the basic model. While, for space reasons, I have left out various (some not insignificant) details, there is enough in the foregoing to turn to philosophical discussion.92

Comments: virtues and strong(-er) truth

I will concentrate on the general semantics of Field's conditional, focusing on the philosophical features. Right off, one of the chief desiderata is plain.

IPA~A

9°1 will continue to use the bar-notation, e.g. IAI, to abbreviate 'the value of A'. (This follows Field's notation, and ties in the earlier discussion of 'ultimate values'.) The difference, of course, is now that values are relative to worlds, and so, e.g., IAl.wis the value of A at w.

91In standard 'non-normal worlds' semantics, wherein one has different 'types' of worlds, broader notions are standardly introduced by various restrictions on the 'types' of worlds invoked in one's definition. Field's neighborhood account is similar, and [Field, 2003b] introduces 'universal validity' (quantifying over all worlds of all interpretations) and 'strongly valid' (all normal worlds of all interpretations). For space reasons, I skip over these distinctions.

92What may not be plain is how Field's 'general semantics' is a more general version of the noted 'restricted semantics'. The short answer is that the latter can be seen as a special case of the former: one allows for 'normal ordinals' (analogous to 'normal worlds') in the latter, and modifies the account of validity in the latter in terms of a distinguished such 'normal ordinal' (something guaranteed by Field's Fundamental Theorem). For discussion see [Field, 2003b].

Hence, given the intersubstitutivity of dT, which is preserved in *all* (transparent) contexts, all dT-biconditionals are similarly valid. And since Field [2003a] gives a consistency proof for the resulting language (and dtruth theory), this amounts to a remarkable step forward in (consistent) paracomplete accounts of dtruth. Kripke provided an answer a paracomplete answer to NTP for a language in which not all dT-biconditionals hold. What Field's conditional has given us is an answer a paracomplete answer to NTP for a language in which all dT-biconditionals hold, an answer that preserves the insights of Kripke but goes further, properly extending the resulting logic.

But there's more. The given conditional also exhibits various familiar features.[93]

A, A---, B I~B
IP ~A ~ A
IP ~(A ~ B) ~ (A V~B)
IP Vxd ~ d(t/x) (fox"proper substitution)

A,~B IP ~(A ~ B)
Ib AAB---~ A
I~ (A ~ ~B) ~ (B ~ ~A)

In addition to other features that are here omitted (for space reasons), Field's conditional exhibits fairly natural behavior, at least within a general paracomplete setting in which LEM is rejected (which, of course, is the aim). In particular, --~ behaves like D when excluded middle is assumed, and so we have, for example,

ADBIPA~B

and, indeed,

A V ~A,B V ~B IP (A D B) ~ (A ~ B)

and similarly for material equivalence.[94]

In the remainder of this (sub-) section I will briefly address a few of the pressing philosophical issues mentioned above: Curry, strong (-er and stronger) truth, and, relatedly, 'exhaustive characterization'.

Curry paradox

Of course, given Field's consistency proof [2003a], one knows that Curry paradox doesn't pose problems; however, it is worth (at least very briefly) touching on the issue and so I will, but only very briefly.

I've assumed, as above, that a minimum requirement on a 'genuine conditional' is that it detach in the sense that the argument from *{A,A ~ B}* to B is valid. Field's conditional is genuine. But Field's conditional also validates Identity, I~ A ~ A, so that (given dtruth) the dT-biconditionals all hold. In such a setting (given normal behavior of other connectives), Curry demands on pain of triviality giving up Reductio in form (R), contraction principles (C1) (C4) and the like, as well as (Imp) and (Int). (See §4.1, page 344ff.) That Field's conditional does as much is not only a virtue but, as said, a necessity on pain of triviality. Since Curry's paradox turns on the validity of such principles or rules, any argument towards Curry paradox is blocked as invalid.[95]

---

[93] For a list of other notable features see [Field, 2003a; Field, 2003b; Field, 2005c], and for a very illuminating discussion see [Yablo, 2003].

[94] Note well: These last features depend on having other features that I've here omitted, features that require other constraints being imposed on the 'neighborhood interpretations'. With suitable constraints imposed, e.g., one gets the 'meta-rule~that if arguments *A/B* and *B/C* are valid, then so too is the argument A V *B/C*. (This does not hold in K3, but does hold in a proper extension of it, sometimes called K3+.) See Field's cited work for a full discussion.

It might be useful to see a 'counterexample' to some of the invalid principles. Consider principle (C1), which is sometimes called *Assertion* or *Pseudo Modus Ponens.* Notice that if @ could be lonely, there would be no way to invalidate (C1). But Field's demand that @ be happy provides an immediate counterexample: just consider an interpretation in which 5c~ = {{~, w}} and, for simplicity, let ~ = 0. Let IAI~ = IBI~ -- IAI~ = 1 and IBl¢,, = 0. Then IA ~ Bl¢,, = 1 and, hence, IA/~ (A ~ B)l~, 5, and so IA n (A ~ B)l~, > IBl~. But most importantly, since not every world in the (unique) @-neighborhood is such that *IAA(A* ~ B) ~ B I < IBI or *IAA(A* ~ B) ~ B I > IBI, we have what paracomplete theorists will naturally want with respect to Curry-instances of (C1).

IA n (A ~ B) ~ BI~ 1

And since 5 is not designated, (C1) is invalid. Sinfilar (and, indeed, even simpler) counterexamples are available for the other principles and rules.

Strong truth: determinacy

Field's aim is not only to validate all dT-biconditionals but also go beyond the Kripkean framework with respect to characterizing Liar-like sentences. Towards that end, Field proposes to recognize a stronger notion of truth than mere (entirely transparent) dtruth.

As mentioned in §4.2, there is always a risk of introducing more 'truth-like' devices (predicates, operators): paradoxical sentences are always ready to spring up. This is where Field's consistency proof [2003a] for the full conditional-ful (and dT-ful) language comes into play.

The consistency proof shows that the conditional doesn't introduce any further paradoxes that aren't already resolved by the guiding, paracomplete rejection of LEM. And that is the key. Field wants to characterize all Liar-sentences via 'stronger and stronger truth', and do so without bringing about yet further paradox. Given the consistency proof, it is natural to seek an account of such 'strong truth' that *invokes only the resources of the language at hand,* the language for which we have Field's consistency proof. A remarkable feature of Field's framework is that he enjoys just such an account: he defines infinitely many 'determinately' operators these amount to 'stronger and stronger truth' *out of the conditional.*

---

[95] I should note that there are various versions of Curry's paradox, not each of which assumes *all* of the 'bad' principles or rules, but each assumes one or more of them. I should also note that, at least in paracomplete settings, one need not give up the *structural* contraction rules, as one does in paraconsistent settings. (Giving up the structural rules prohibits, in effect, 'using a premise more than once' in a proof. Field's framework preserves structural contraction, so understood.)

There are two items that need to be explained (or, at least, sketched): what is the definition of 'determinately' in terms of the conditional? How do we get infinitely many? I will (very) briefly sketch the answer to each question, leaving details to Field's work (cited throughout).

The basic account. Where • is any logical truth, Field proposes to 'introduce' *a determinately* operator D thus:

DA aS (T~A) A A

This immediately gives standard behavior for 'determinately' operators, in particular,

dl. A Ib DA

d2. DA Ib A

d3. IP DA-~ A

As mentioned in §4.2, standard approaches to 'determinate truth' usually give up the converse of (d3), and the situation is no different with Field. That the converse does fail may be seen by considering an interpretation in which *IAIc~* ·1, in which case IT -~ *AIc~* is not designated, regardless of ~c~.

With D at hand, one can now classify D-free Liars as proposed: neither they nor their negations are *determinately* true. For any such sentence A, we may assert ~DdT(A} A ~D~dT(A} or, equivalently given dtruth, ~DA A ~D~A. 9s

But what of weaker Liars? Now that we have D, so understood, in the language, we inevitably get D-ful Liars for example, sentences that say of themselves (only) that they're not determinately true, or not determinately determinately true, or so on. Such sentences call for 'stronger and stronger truth'. But Field's construction already yields as much.

Infinitdy many. The point is fairly obvious. Consider a Liar L that says ~DL. That Field's construction handles L, so understood, follows from his consistency proof: such a sentence receives an interpretation in the language, namely, 5" Of course, given (dl) (d3), one cannot truly classify L as being not determinately &rue. But one can generalize Field's proposal in a natural way: one can truly classify L as *not determinately determinately dtrue,* that is, one may assert ~DDL.

In general, Field's operator may be iterated into the transfinite: for some suitable ordinal notation that yields or, we have an operator D~, the or-many iteration of

---

96Recall that Field's construction preserves dtruth, that is, preserves the essential intersubstitutivity of dT(A} and A in *all* (transparent) contexts, including 'inside' conditionals and, hence, qnside' D-contexts.

D. And for each such 'determinately' operator, there will be (increasingly weaker) D~-ful Liars, each of which gets 'correctly characterized' by a stronger operator II}~+1 (provided the ordinal notation yields as much).[97]

What is central to the proposal is its thoroughgoing paracompleteness. LEM does not hold even for 'determinate truth'. Consistency is purchased by such thoroughgoing paracompleteness: for any D~-ful Liar L~ (for suitable 0), the failure of II}~L~ V ~II}~L~ will arise. Of course, as mentioned in §4.2, the failure of LEM for one's level-or 'strong truth device' requires having an even stronger device if one wants to classify level-~ Liars. But the point is that Field's basic construction provides as much. For any Liar constructible in the language (or, at least, the hierarchy of determinately operators), there's a 'strong' oi" 'determinate' device that classifies the sentence.

And to repeat: there is no threat of 'determinately'-Liars wreaking havoc, since all such Liars are constructible only in a language (the full conditional-ful language) that enjoys a consistency proof. Needless to say, this is a significant improvement on the Kripkean picture.

Exhaustive characterization ?

Here, the issue is delicate. What Field provides is a way of consistently classifying all Liar-like sentences (would-be paradoxical sentences) definable in the 'hierarchy of determinately operators'. This is a major step forward on other paracomplete accounts. On the other hand, one might wonder whether Field has achieved 'exhaustive characterization' in the target (but, alas, as yet imprecise) sense. Can we, for example, dtruly assert that every sentence is dtrue, dfalse, or not determinately dtrue in any (of the infinite) sense(s) of 'determinately'? Better (but still not ideally) put, is there some unified 'semantically significant' predicate Other such that we can dtruly say the following?

dT<A} V dF<A} V Other(A}

The answer is No, at least if Other carries the intuitive sense of *not determinately dtrue or determinately not dtrue, in some sense of 'determinately' or" other'*. In that respect, Field's proposal, while a significant advance on paracomplete approaches, seems to fail to achieve the target sense of 'exhaustive characterization' involved in Ecp, imprecise as that target sense remains.

But the issue, as mentioned, is delicate. Such a 'characterization' presupposes the existence of some 'absolute determinacy operator' that, in effect, would be (like) the union of all determinacy operators. Given Field's consistency proof, such an operator simply doesn't exist, at least if our 'real language' is relevantly like Field's model language. One might, of course, say that Field's model is inadequate precisely on that score: our 'real language', unlike Field's model language, has the resources to express just such a unified ~determinacy' operator, one that is entirely intelligible. But Field [2005b] questions such alleged intelligibility.

---

[97] For limit ordinals, one can mimic infinite conjunctions via the dtruth predicate and a suitable ordinal notation.

Unfortunately, Field's position on the current matter is (well) beyond the scope of this essay, but the basic issue can be (very roughly) sketched as follows. What is remarkable about Field's proposal is that every sentence in the 'hierarchy of operators', and hence any Liar sentence in the given hierarchy, gets characterized by some (semantically significant) predicate or other in particular, some predicate constructed out of 'stronger truth' (determinacy), negation, and so on. Call such predicates *indeterminacy predicates.* It would seem that one could *conjoin* all such indeterminacy predicates to get one %ig, unified indeterminacy predicate', or at least get the same effect of such 'conjunction' by quantification and dtruth. All of this seems to be eminently intelligible. The trouble, according to Field, is that the apparent intelligibility is merely apparent.

In short, Field [2005b] shows that such a 'conjunction', to the extent that it is achievable, either cannot be (coherently) constructed or, to the extent that it can, won't behave as expected; the reason, in keeping with a thoroughgoing paracomplete account, is that it is ~indeterminate' what the conjuncts are! Any such quantification, aimed at achieving the (alleged) unified indeterminacy predicate, would at best be restricted quantification involving an ~indeterminate' or ~fuzzy' restricting condition.

The sense in which such 'conjuncts' are ~indeterminate', or the restricting condition 'fuzzy', is quite involved and, unfortunately, too involved for this essay. But it is worth noting that Field has a reply to the charge that he fails to achieve the target sense of exhaustive characterization. For a full, detailed discussion, see [Field, 2005b].[9s]

~.3 Paranormal

Kripke showed how we can have a consistent (and, hence, nontrivial) dtruth predicate, but not for a language in which all dT-biconditionals hold or for a language in which we can dtruly characterize Liar-like sentences (let alone 'exhaustively' so characterize). Field, in turn, show's how we can have a consistent dtruth predicate in a language that validates all dT-biconditionals and enjoys the resources to characterize all Liars constructible in the language (or, at least, in the 'hierarchy of determinacy operators'). What Field does not and, in his framework, cannot consistently do is offer a 'unified predicate' that characterizes all Liar-like sentences. Deviating from Field's framework, this section (briefly) sketches a route one might take towards achieving a unified 'classification' of Liars.[99]

[9s] While Field's proposal is clearly an advance on earlier paracomplete accounts, I should note that questions remain open as to whether Field's proposal sufficiently covers related semantical paradoxes, in particular, paradoxes of denotation (e.g., Berry's). Graham Priest [2005] argues that Field can handle such paradoxes, if at all, only via ad hoc restrictions. Field [2005a] disagrees, maintaining, as any paracomplete theorist will, that Berry's paradox is problematic only if some illicit use of LEM is invoked. But the issue is tricky and, alas, beyond the scope of this essay.

Paracomplete theorists recognize the need to expand our otherwise classical categories. We have dtrue sentences and, derivatively, dfalse sentences; and there are 'other' sentences 'in addition to the normal ones', sentences like Liars. Call such sentences *paranormal,* again from the Greek 'para'. How', if at all, can we achieve a consistent dtruth predicate, all dT-biconditionals, and also 'exhaustive characterization' in the target (but, admittedly, imprecise) sense, one for which we have a unified 'paranormal' predicate? I will briefly sketch an answer, one that, at least philosophically, depends on a particular conception of the *role* of (the predicate) 'paranormal'. I will not discuss a suitable conditional, but it should be clear that something along the lines of Field's 'neighborhood' conditional will be available.[100]

Philosophical picture

What the Liar (and its ilk) teaches us is that besides the dtrue sentences and dfalse sentences, there are paranormal sentences. One might now wonder: what is it to be a 'paranormal' sentence?

The suggestion is that we set the question aside. For present purposes, it suffices merely to 'tag' the target sentences (e.g., Liars) as such, namely, as *paranormal.* Ultimately, there may well be no interesting property of *being paranormal,* and accordingly no hope of informative 'analysis' or explication of 'paranormal'. But the term may none the less serve to give us the sort of 'exhaustive characterization' desired, just by giving us a 'logical device' of sorts with which to 'classify' the target sentences. My suggestion is that we resist questions concerning 'the nature' of paranormals, seeing it merely as a tag (a logical category) introduced for the target sentences.

Notice that even at this stage without giving much more than a 'classifying role' for the device Liar phenomena already arise.

✓ The ticked sentence in §4.3 is either not dtrue or paranormal.

And such a sentence itself is surely among the very sort for which we introduced the tag 'paranormal', and indeed the usual Liar-reasoning will suggest as much given the relevant version of Ec: namely, that every sentence (hence, the ticked one) is dtrue, dfalse (true negation), or paranormal.

The upshot is that *paranormality and dtruth apparently overlap.* And my suggestion is that we simply accept as much. After all, we want a simple, exhaustive characterization, one that is consistent. And we can enjoy as much by acknowledging that some paranormal sentences just in virtue of the role of 'paranormal' and the basic expressive job of 'dtrue' turn out to be dtrue.

---

[99] This section is drawn from [Beall, 2006].

[100] Indeed, I'm inclined to think that an even simpler conditional, invoking so-called non-normal worlds, will do the trick. For a sketch of such a conditional albeit in a paraconsistent setting see §5.1.

Likewise, of course, various paranormal sentences will inevitably be dfalse, for example,

* The starred sentence in §4.3 is not paranormal.

If we have it via the relevant version of EC that the starred sentence is dtrue, dfalse, or paranormal, then it's dtrue iff not paranormal (and, so, not dtrue and paranormal). So, the sentence (just reasoning intuitively, at the moment) is paranormal, and so... a dfalse paranormal.

One might picture the story as follows, although one ought to keep in mind that this is only a heuristic.[101]

One might press for analysis or explication: what is it to *be* paranormal?! The suggestion is that we resist the question. Truth itself (at least on a suitably deflationary conception) affords little by way of informative analysis. On the usual picture, we began with our dT-free fragment and had no problems except expressive ones due to our finite limitations. We could neither implicitly nor explicitly assert everything that we wanted to assert. Towards that end, our 'dtruth'-device was introduced. But once 'dtrue' was introduced (into the grammatical environment of English), various unintended sentences emerged typical Liars and so on. Towards 'classifying' *those* sentences, 'paranormal' is introduced. But given the job of 'paranormal', there's little reason to expect let alone demand an informative analysis. Indeed, as is evident, there is even less to say about 'paranormal' than 'dtrue'.[102]

Notice that by allowing 'overlap' between the paranormal and the dtruths, we thereby avoid the need to invoke infinitely many (non-unifiable) 'stronger truth' predicates. Once 'paranormal' is introduced, unintended by-products of it emerge *this sentence is paranormal,* etc. The suggestion is that we simply let such sentences be among the paranormal, even though given the role and rules of 'dtrue' they may likewise be dtrue. (Similarly for 'dfalse'.) If our chief concern is to 'exhaustively characterize' or 'classify' in a *consistent* and *simple* way, then such overlap is harmless, provided that 'dtrue' and 'dfalse' avoid overlap.

---

[101] And see §4.3 for an alternative (and, in my view, more attractive) picture.

[102] But see §4.3 for an alternative picture.

Notice, too, that the proposal is *not* motivated by a search for some suitably stronger notion of 'truth' with which we can assert the 'untruth' (in a stronger sense of 'true') of Liars. The proposal is simply that we classify such sentences as, well, paranormal and that is that. If more is required for serious inquiry (e.g., science), then the proposal is wanting in that respect. But, as far as I know, not more is required. Moreover, one can, of course, define a stronger notion of truth via dtruth and 'paranormal' if one wishes. (See *Strong truth*, page 366.) But doing so seems not to be a requirement.

There may be philosophical issues to be sorted out, but they must wait for another time. For now, I briefly turn to a simple, formal sketch of the idea.

A formal model

The picture is along familiar many-valued lines.[103] Our 'semantic values' (in the *formal story)* are elements of V = {1, $\frac{3}{4}$, $\frac{1}{2}$, $\frac{1}{4}$, 0}, with designated elements in ~) {1, $\frac{3}{4}$}.[104]

Our atomics are interpreted via a function u in the usual way,[105] extended to compounds along the lines seen in Strong Kleene: u(AAB) is the minimum of u(A) and u(B), and u(A V B) the maximum. (Quantifiers can be treated similarly, as generalised conjunction and disjunction.) Negation is likewise familiar: u(~A) = 1 - u(A). Hence, negation is fixed at 1 but otherwise toggles designated and undesignated values; it is thus 'normal' in the usual (formal) sense.

We assume a special predicate dT to be interpreted as a dtruth predicate: u(dT(A)) = u(A) for any 'admissible' u. Falsity, in turn, is derivative: u(dF(A)) = ~(dT(~A)).

Finally, we add a unary connective 7c (our 'paranormal' device), which is interpreted thus:[106]

$$u(TcA) \quad \begin{cases} 0 & \text{if } u(A) \ c\{1,0\} \\ ] & \text{otherwise.} \end{cases}$$

Note that letting 'P' be our 'paranormal' predicate the extension of P, namely,

---

[103] NOTE WELL: While I do not (here) discuss a suitable conditional, the conditional in question is *not* to be constructed along standard Lukasiewicz lines (for reasons mentioned in §4.1.). The conditional will be one that is not 'truth-functional' but, rather, either invokes 'non-normal points' familiar from Kripke and, more recently, work in 'relevant' semantics (see §5.1 for brief sketch), or something along the lines of Field's 'neighborhood'.

[104] Logical consequence semantic validity is defined as usual in terms of ~D. I skip the definition here.

[105] Note that if, for some reason, one wanted to have a three-valued language, letting)2 contain only three values, one could let interpretations be non-functional relations and given appropriate constraints on the interpretations thereby achieve the same framework. Functional interpretations are more familiar, and so I go with that.

[106] Note that I will use '~r' for both the connective and operator, trusting that context will do its clarifying job.

$$79+ \quad \{_{A:\sim} 1_{<p(A)<} \quad 3\}$$

may well be negation-inconsistent; it may well contain both A and ~A for some A. (Indeed, for typical 7c-free Liars, that will be the case.) In this respect, *being paranormal* differs from *dtruth* (similarly, *dfalsity)*, as the extension of the latter, namely,

$$T+ \quad \{3_{A:\ \sim\ <\ \sim(A)}\}$$

is always negation-consistent.[1°7]

Comments

What the 'paranormal' approach yields is a simple, consistent way towards achieving exhaustive characterization in the target sense, one that employs a unified 'paranormal' predicate. Liars, one and all, may safely be classified as paranormal, even though some paranormals may also be dtrue or dfalse. And, in general, we achieve 'unified exhaustive characterization', in that every sentence will fall under one of our three semantically significant predicates: either dtrue, dfalse, or paranormal. In this section I briefly address two issues, leaving further discussion to other work.

An alternative picture: residual

As above, the 'paranormal' account achieves a simple, unified exhaustive characterization (while, suitably filled out, preserving the salient virtues of other paracomplete accounts). Of course, such exhaustion such 'unification', as it were comes at the 'cost' of there being nothing illuminating to say about *being paranorreal.* How much of a cost that may be is not entirely clear. Still, one might at least like to have it that, for example, if a sentence is not dtrue, then it is paranormal, and if not paranormal, dtrue. We do not have as much on the foregoing account, but one can easily tweak the picture to get as much.

One might, for example, settle on a (fairly common) thought according to which we have our *dtrue* sentences and, beyond that, we simply have *the rest* be they dfalse or whathaveyou. Here, the picture is one according to which our goal (e.g., in science or rational inquiry, generally) is to record the full, dtrue story of our world and simply chuck the 'remainder' to the side. On this picture, anything *dfalse* (having true negation) is 'residual' and, moreover, anything *not residual* dtrue. The picture might look as follows, with dfalsity now merely a proper part of the residual.

[1°7] A consistency proof cannot be done in the usual fashion, given that ~r is nonmonotonic, but such a proof is available. (Thanks to Tim Bays and Greg Restall for their interest in the idea; each of them independently suggested different proof-sketches.)

dT R

Such an idea requires only a minimal tweak of the 'paranormal' account. In particular, instead of 'paranormal', define our 'residual' tag thus: l°s

0 if ~(A) = 1
v(t A) 3 otherwise.

In turn, the predicate 'R' (residual) behaves thus:

7-4+ {A : 0 < p(A) < 34}

7~+ will differ from 'dfalse' by being (negation-) inconsistent, whereas the extension of 'dfalse' will be (negation-) consistent:

5 + {A : 0 < ~(A) < 14}

This variation reflects the 'residual' conception according to which any dfalse sentence is residual, and any sentence that is not residual is dtrue. The salient difference between 7c and t is that we have

~A Ik-tA

and

~tA Ik A

only for t, not for 7c.1°9 This reflects the idea, as above, that (in some sense) science or, generally, rational inquiry aims to separate dtruth from the 'remainder', the residual. But, as before, once we add (even a minimalistically construed) category 'residual', Ec will demand overlap, and so we allow that sortie of the resulting residuals are true. But no matter; the given overlap is harmless, and also allows for a simple 'unified exhaustive characterization'.

108To avoid confusion, I will use 't' for our target 'residual' tag (rather than simply redefine ~7l-~).

l°9while I am not discussing a suitable conditional here, the (target) conditional versions of the above principles will similarly hold.

Strong truth

Let me make it plain that the foregoing account is not motivated by a desire to say that Liars are in some sense 'not true'. Still, one can say as much if one wants. For example, on the 'paranormal' account (but similarly for 'residual'), define *robustly true* thus:

TA iff *A A~rcA*

Obviously, we have it that t,(qFd) 1 iff t,(d) 1. As above, t,(rcA) c {3 0} and, hence, ~(~rcA) E {1,1} for all ~.

We immediately get principles reminiscent *but atypical* of standard 'determinately' operators.

* *TAIbA.*

* *A~ TA. (Just let ~(A) = ~.)*

The second principle is what is atypical with respect to standard *'determinately'* operators, especially since A ~ TA will likewise fail, at least on the intended construction (about which, for space reasons, I've said little here). While I won't try to defend such 'failures', it seems to me to make sense: we are acknowledging that some paranormals may be dtrue (true in our basic, merely expressive, entirely transparent sense) but not thereby 'robustly true', and that this point carries all the way through both for the 'validity' reading and the conditional reading, n°

The advantage (if any) of having such 'robust truth' is that it may do some work that our fundamental dtruth predicate isn't cut out to do. Our fundamental dtruth predicate has only the job of being a transparent generalization device, of being such that dT(A} and A are intersubstitutable for any A, including, of course, the paranormal A. The value (if any) of 'robust truth' is that it may cut distinctions that dtruth itself can't and was never intended to cut. In particular, one can dtruly say that, for example, the ticked sentence in §3 is not robustly true. Of course, what, if anything, one gains from this over and above merely asserting that the given sentence and its negation are 'residual' (or paranormal) is not obvious. But, for space reasons, I will leave the matter there. 111

5 PARACONSISTENT

'Para' here is used as in 'paracomplete' (see §4), and 'consistent' for negation consistency, the idea being that we are moving 'beyond' typical (negation-) con-

---

110That these 'readings' inevitably come apart in paracomplete (and paraconsistent) approaches is briefly discussed in §6.

111I will note one more thing. T, defined as above from our 'paranormal' (similarly 'residual') device, seems to behave very much like if not exactly like the KF-truth predicate. This is philosophically interesting, as it gives us the (e.g., expressive) virtues of dtruth but likewise a stronger notion of truth that has gained independent support from others. But details are left for elsewhere.

sistency constraints.[112] What is common to paracomplete theories, as in §4, is a rejection of LEM. What is common to *paraconsistent* theories, in turn, is a rejection of EFQ (ex falso quodlibet or, more colorfully, explosion), the rule according to which (arbitrary) B follows from (arbitrary) A A ~A or from {A, ~A}.[113]

A paraconsistent logic, then, is one in which EFQ fails: A, ~A N B. An *explosive logic* is one in which EFQ holds. Any theory hence, any dtruth theory according to which both A and ~A are dtrue is *trivial* if its underlying logic is explosive. Paraconsistent logics afford nontrivial but (negation-) inconsistent theories. And that's the basic idea behind paraconsistent dtruth theories, in general.

While paraconsistent approaches to truth are far from dominant (to say the least), there have been various proposals either directly in the tradition or very close to the spirit of it, including [Dowden, 1984], [Priest, 1979; Priest, 1987], [Rescher and Brandom, 19791, [Shaw-Kwei, 19541, [Visser, 19841, [Woodruff, 19841, variations of [Yablo, 1993a; Yablo, 1993b], and others.[114] For present purposes, I will focus on Priest's so-called dialetheic account.[115]

I should note that in paraconsistent theories, as in their paracomplete relatives, 'reject A' is not equivalent to 'accept ~A', and probability requires modification (in effect, the dual of the paracomplete modification). For discussion of (minor) adjustments to probability theory, see [Priest, 1987].

5.1 Priest

Graham Priest takes both NTP and ECP seriously, and, in a large body of work, argues that only by acknowledging 'gluts', sentences that are both dtrue and dfalse, do we achieve satisfactory answers to those projects,[116] Priest (and Richard Sylvan, formerly Routley) coined the term 'dialetheism' *(die-a-lee-thism)* for the view that some sentences are both dtrue and dfalse equivalently (in the given frameworks), that some dtrue sentences have dtrue negations. Whether dialetheism is the only satisfactory approach to NTP and ECP is something that I leave open or, at least, to (arguments in) Priest's cited work. My aim, as throughout, is only to sketch the basic idea.

---

[112] Actually, paraconsistentists seem to have different views of the import of 'para', but I will ignore this here. For discussion, see [Priest, 2000].

[113] Some well-known paraconsistent logics are such that 'conjunction' is abnormal either Simplification or Adjunction fails. See [Priest, 2002] for broad discussion. I will skip discussion of such logics and focus entirely on 'normal' conjunction.

[114] Martin, Dowden, and Yablo all (independently) tweak Kripke's iterative proposal to achieve an 'inconsistency' account. (Also, Field [2005a] briefly discusses a paraconsistent 'dual' of his own iterative construction.) Rescher Brandonfs and Priest's accounts are compatible with such an iterative construction, but they instead give a simple, non-iterative account. Visser's given discussion is an excellent discussion of fixed point technicalities in four-valued frameworks.

[115] priest [1987] does not actually propose an account of *dtruth,* and instead imposes various restrictions on truth to avoid full intersubstitutivity. None the less, I will present the basic idea in terms of dtruth.

[116] priest uses the terms 'dialetheia' and 'true contradictions' for *gluts.* I will use 'gluts', which is a term coined by Kit Fine [1975].

I should note one caveat: that *dialetheism* and *paraconsistency* are not one and the same! While all paraconsistent logicians see paraconsistent logic as a useful tool for modeling inconsistent but nontrivial theories (e.g., naive truth, naive semantical properties, etc.), many if not most reject that such theories are 'possibly dtrue'. Dialetheists are a small minority, holding that some of the given theories are not only possibly dtrue, but actually so.[117]

Philosophical picture

According to dialetheism, what the Liar paradox teaches us is that some dtruths have dtrue negations, specifically, Liars.[118] The lesson, I think, is quite natural. After all, 'dtrue', as the (typical) story goes, was introduced solely as an expressive, entirely transparent device: for *any* sentence A, be it dT-ful or otherwise, dT(A} and A are intersubstitutable. But, of course, once the device (a unary predicate) is introduced into the language, Liar-like sentences emerge, sentences that say of themselves only that (for example) they are not dtrue, that they are dfalse. Given LEM, such sentences turn out to be dtrue and dfalse.

The *paracomplete* response, as above, is to reject LEM, preserving the (negation) consistency of dtruth. The *dialetheic* response is to accept the inconsistency of dtruth but reject EFQ, thereby preserving the nontriviality of dtruth. Both approaches agree that our 'classical categories' need to be expanded there's something 'in addition to' the normal categories. The dialetheist's additional category is *glutty*, that is, *both dtrue and dfalse*, a category characterized (in the language itself) just so, 'is dtrue and dfalse'.[119]

Notice that rejecting EFQ is not ad hoc. To begin, why think that (arbitrary) B follows from (arbitrary) A and ~A? One answer might be that there's no apparent counterexample. Indeed, one (many?) might think it 'inconceivable' that both A and ~A be dtrue, and hence that, vacuously, the given inference count as valid. Such an answer, I think, is entirely natural and, in general, quite reasonable; however, it is not ultimately sufficient. On one hand, intuitions are 'built' from normal, run-of-the-mill cases usually cases 'grounded' in non-semantic facts (in a minimal sense of 'facts') and, in such cases, it is difficult (to say the least) to understand what it would be like for $A \wedge {\sim}A$ to hold.[120] On the other hand, EFQ is a principle about *all* cases (all sentences), not just the run-of-themills. When one takes Liars into consideration, especially with respect to (entirely transparent) dtruth, it is fairly easy to see how for some A, namely, an A that says of itself (only) that it is dfalse, both A and ~A could be dtrue. (Just consider the intuitively correct reasoning of the Liar paradox!)

---

[117] For discussion of different 'grades of paraconsistency', see [Priest, 2000]. (And for a reply to Priest's alleged 'slippery slope' towards dialetheism, see [Beall and Restall, 2005, Part IV].)

[118] Note that whether a given Curry sentence is a Liar depends on the conditional. If the conditional is the hook (material conditional), then dialetheists (at least those, like Priest, who accept LEM) will count the sentence among the gluts. (Such a sentence will be nothing but a familiar disjunctive Liar 'I am not dtrue or everything is dtrue'.) If a genuine (detachable) conditional is at work, then dialetheists on pain of triviality will reject the sentence. See §5.1 for brief discussion of a suitable conditional.

[119] I will briefly address some 'characterization' worries in §5.1.

[120] But, alas, Priest [1987] argues that even in non-semantic, sometimes run-of-the-mill cases, there is good reason to accept $A \wedge {\sim}A$ (for some A). I will ignore such considerations here, for space reasons.

Another reason one might give for the validity of EFQ invokes ~the very meaning of negation'. One might say that what we *mean* by negation is such that EFQ holds. But this reply either begs the question or is confused. After all, part of the job of a logical theory is to 'describe' the behavior of our various so-called logical connectives, including negation. The question at hand is whether EFQ holds. To say that EFQ holds in virtue of the very meaning of negation is to either beg the question or confuse theory with subject matter.[121] Given the role of dtruth and the existence of Liars, it is not obviously unreasonable to conclude, as the dialetheist does, that some sentences are both dtrue and dfalse, and hence that some A is such that A A ~A is dtrue.

That rejecting EFQ is not ad hoc is now plain. If one takes seriously the idea that a sentence say, a Liar could be both dtrue and dfalse, then arguments for the validity of EFQ are hard to find. C. I. Lewis' famous ~independent argument' for EFQ [1932] is faulty precisely for its not taking the initial supposition seriously. After all, (seriously) suppose that A and ~A are (both) dtrue, in which case, presumably, A V B is dtrue, since, as above, at least one of the disjuncts is dtrue (viz., A). But now the point: there is no reason at all to think that B follows from the (supposed) dtruth of ~A and A V B. Indeed, it is easy to see why this inference namely, Disjunctive Syllogism would fail, at least if both A and ~A could be dtrue.[192]

What about ECP? Does the dialetheist face the predicament of paracomplete theorists (e.g., Field) who, towards characterizing all Liar-like sentences, are forced to acknowledge ~stronger and stronger truth'? No, at least not obviously. (But see §5.1 for some muted worries.) The paracomplete theorist's trouble emerges from rejecting LEM, and thereby losing the means by which to classify Liars (for which LEM fails). (This is particularly salient in Kripke's proposal, as above.) But the dialetheist need not reject LEM. Indeed, Priest [1987] argues that once EFQ is rejected (on the basis of recognizing gluts), there is no good argument for rejecting LEM.[193] Moreover, and perhaps more centrally, the dialetheist essentially recognizes overlap among her semantically significant categories. Such overlap, as in the ~paranormal' approach, affords a unified predicate in terms of which all Liars may be classified.

---

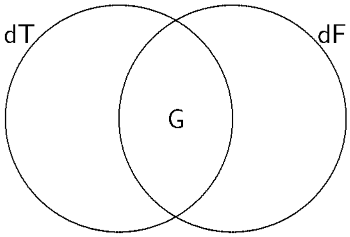
[121] Quine,s famous 'dilemma of the deviant logician' seems to make just such a confusion. See [Quine, 1970].

[122] For more on 'the very idea of gluts', and related issues, see [Priest, 1998].

[12a] I will briefly return to this issue in §5.2, wherein a ~gappy and glutty' (dialetheie) position is (very briefly) suggested.

Formal picture

Priest's basic picture, as above, is one according to which our *Other* category *glutty* overlaps both dtrue and dfalse; indeed, the additional category just is the given overlap (with G for *glut@)*. A picture is as follows.



This picture is the 'dual' of the Strong Kleene picture, which likewise has three semantically significant categories, but, unlike the above, none of the Strong Kleene categories overlap. Indeed, Priest's purely extensional *LP* [1979] just is the dual of/(3, and so clauses for conjunction (minimum), disjunction (maximum), and negation remain exactly as before, namely (K1) (KS). (See §4.1.)

LP, then, is exactly like/(3 except that the 'middle value' is designated. Designating the middle value, which represents 'both dtrue and dfalse', reflects the idea that any dtrue sentence even if also dfalse is assertible. Letting T + and T- be the extension and antiextension of dT, and letting $ comprise all sentences (in the augmented dT-ful language), the chief contrast between/(3 and *LP*, at least semantically, comes to this:

K3. Every interpretation is such that T + A T 0 but some interpretations are such that T + U T- ¢ 8.

LP. Every interpretation is such that 7-+ U T- = $ but some interpretations are such that T + A T ¢ 0.

So, while paracomplete approaches put Liars 'beyond' T + U T , Priest's (dialetheic) paraconsistent approach puts Liars into T + A T .

Validity is defined exactly as in the Ka framework: E I~-A iff every interpretation that designates (each element of) E also designates A. In other words, E IF-A is valid iff for all interpretations p and all B c E, if p(B) c {1, 1}, then p(A) c {1, 1}.

With validity so defined, it is immediately clear that *LP* is paraconsistent, that is, that A, ~A P5/3. Just consider an interpretation such that ~(A) = 1 = ·(~A) but ~(B) = 0.

Similarly, the above counterexample to EFQ also shows that disjunctive syllogism (DS) is invalid in *LP: A V t3, ~A ~z t3.* This 'failure' is especially significant inasmuch as DS is equivalent to 'material modus ponens', the rule that B follows from A D B and A, where *A D B,* as before, is defined ~A V B. But, then, *LP* fails to have a genuine conditional, an issue to which I return in §5.1.

So goes the basic (extensional) framework. The resulting account is one that answers NTP by rejecting EFQ, and answers ECP by recognizing a unified predicate (say, 'glutty') applying to all Liar-like sentences. I turn to a few issues, including the issue of a suitable conditional.

Comments

In this section I address only a few philosophical issues that arise with respect to Priest's dialetheic approach to dtruth.[124]

Non-triviality project

Recall that NTP is the project of explaining how we can enjoy a nontrivial language that has a dtruth predicate and Liar-like sentences. In Priest's case, the project is not to show how we can enjoy a *consistent* dtruth predicate despite Liars, but how we can enjoy nontriviality. As in §4.1, one generally aims to answer NTP by constructing an artificial, formal language the *model language* that contains its own dtruth predicate, and then claims that, at least in relevant respects, 'real dtruth' is modeled by truth-in-the-model language.

Just as in the Kripke case, one might object that Priest's account doesn't sufficiently answer NTP, the reason being that certain notions used in the 'metalanguage' are not expressible in the object- or 'model language'. In particular, one might argue just as in the Kripke case: an LP-based account has it that some sentences are 'both true and false', some 'simply true', and some 'simply false'. Let £.~ be a typical LP-based 'model-language'. In a classical metalanguage, one defines '£.~-simply true', '£,~-simply false', and, of course, 'L:,~-both-true-and-false'. Since such notions are defined via classical set theory, one can, in turn sticking within the (classical) metalanguage prove that L:.~ cannot define, for example, {B : B is £.~-simply false}. The charge, then, is that, on pain of triviality (versus 'mere negation-inconsistency'), £~r~ fails to be an adequate model of real dtruth, since £.~-truth achieves nontriviality in virtue of lacking (the given) notions that we have expressed in our real language.

So put, the charge faces the same problem as the related Kripke case: it is either confused or unwarranted. First, notice that classical logic is an extension of the logic *LP*. Semantically, every classical interpretation is an *LP* interpretation, and thus Priest's model language can enjoy a proper, classical fragment. The idea, of course, is that we are classically modeling our 'real, nonclassical language'. But, then, notions defined squarely within the classical metalanguage a proper fragment of one's real language are merely model-relative ones that, not surprisingly, behave entirely classically. What one needs to show is that there are non-model-relative notions that are expressed in our real language but not expressible in the model language. Pending some (as yet not given) reason to think that the real language has some non-model-relative 'untruth predicate' that behaves over our whole (real) language as our model-relative '/:.~-simply false' behaves over/2~, the given charge is either unwarranted or simply curious, given the aim of £~.

---

[124] Actually, once again, I should note that Priest doesn't himself give an account of *dtruth,* but his framework, as above, obviously affords such an account. See [Priest, 1987] for Priest's preferred account of truth.

Exhaustive characterization ?

One might think that the dialetheist does not achieve exhaustive characterization in the target sense.[125] After all, the dialetheist maintains that all sentences are either *just dtrue, just dfalse,* or *both dtrue and dfalse.* But what predicate does the dialetheist use to express 'just dtrue'? In the formal story, 'just dtrue' is modeled by 1, 'just dfalse' by 0, and 'both' modeled by $\frac{1}{5}$" But, again, how is 'just dtrue' or 'just dfalse' expressed in the real language? If 'just dtrue' is a semantically significant category *distinct firm* 'dtrue', then the dialetheist needs an account of it. Unfortunately, no obvious candidate emerges. (One might think of something along the lines of ~(A A ~A), but this is valid in *LP,* and hence cuts no distinctions.)

Another way putting the worry is as follows. In order to characterize simple dT-ful Liars, standard paracomplete theorists invoke some stronger notion of truth, say ~. But, now, in order to characterize ~2-ful Liars, such theorists invoke an even stronger notion of truth, say ~2. And so on. But now consider Priest's dialetheic approach. While there is no problem characterizing Liars they are both dtrue and dfalse there is an apparent problem characterizing the 'normal sentences', the 'non-dialetheia', the sentences that are 'just dtrue' or 'just dfalse'. The problem seems to be exactly analogous to the standard paracomplete theorist's problem. Whereas the standard paracomplete theorist is pushed towards a stratified conception of 'strong truth', the dialetheist, at least on the surface, seems to be pushed towards a stratified conception of 'just dtrue'.[126]

I am not sure what dialetheists should say about this (admittedly, as yet vaguely sketched) worry. Priest [1987] has long maintained that 'just dtrue' is as inconsistent as dtruth itself, a claim that is neither unreasonable nor surprising. (Consider a sentence like 'this sentence is just dfalse' or the like.) But such a response does not obviously get to the heart of the worry. The point of rejecting EFQ is to allow *some* sentences to be *both dtrue and dfalse* while still enjoying (many) sentences that are 'just dtrue'. This central claim of dialetheism utilizes what one would take to be dialetheism's key semantic categories *both dtrue and dfalse,* and *just dtrue.* But if that is right, and if 'just dtrue' is supposed to be distinct from 'dtrue'

---

[125] The following issue is related to discussion in both [Parsons, 1990] and [Shapiro, 2004].

[126] I should note too that in Field's framework, just such a stratified notion of 'just dtrue' is available to the dialetheist. (Of course, one would in effect dualise the Field construction.) In particular, as c~ increases, DaA rules out more and more of the gluts, the 'dialetheia'. See [Field, 2005a] for some discussion.

(as it is in the formal model, where only the latter is represented by all designated values), then the dialetheist doesn't obviously achieve full characterization in the target sense. But I will leave the matter there.[197]

A suitable conditional

With respect to a conditional, Priest's proposal notably contrasts with the K3 Kripkean proposal (§4.1). While Kripke's proposal enjoys a genuine (i.e., detachable) conditional, it fails to enjoy a conditional that validates all dT-biconditionals. Priest's *LP*, on the other hand, enjoys a conditional that validates all dT-biconditionals, since IP A D A in *LP;* however, the conditional isn't genuine, since it fails to detach.

Unlike Kripke, Priest (like Field) takes the issue of a suitable conditional seriously, where a 'suitable conditional' is a *genuine* conditional that avoids Curry (e.g., Contraction, etc.) and validates all dT-biconditionals. While Priest has suggested various accounts, I will briefly indicate a particularly simple one [Priest, 1992].[12s]

Since the basic issue concerns Curry and detachment, I will here ignore negation. (As will be evident, the framework affords various approaches to negation, including a natural LP-approach.) We expand the language along modal lines, invoking points of evaluation worlds. Exactly how this is done is not pressing, for present purposes. To make matters simple, we will take a 'propositional' approach, letting I I be a function from atomic sentences into p(W), the idea being that **IAI** is the 'proposition' expressed by A, the set of worlds at which A is true. Our set of points W is the union of two sets, iV (normal points) and iViV (non-normal points),[199] with a distinguished element ~ E iV (the actual world) and iV O A/iV 0. With respect to extensional connectives (here ignoring negation), I I is expanded as one would expect.

IA n BI = IAI n IBI

and

I vBI IAIu **IBI**

Finally, interpretations come equipped with an 'arbitrary evaluator' ~ the task of which is to assign values to ~-claims at non-normal points: ~ is a function

---

[127] A related but, in my opinion, not terribly troubling issue is that, for example, an LP-based approach to dtruth is committed to saying of Liars that they are neither dtrue nor dfalse; for, such LP-based dialetheists assert A A ~A for any Liar A, but that is equivalent in *LP* to ~(A V ~A), which, given intersubstitutivity, would amount to the claim that A is neither dtrne nor dfalse. See [Field, 2005a] for discussion.

[12S] For Priest's latest thoughts on a suitable conditional, see [Priest, 2006a; Priest, 2006b]. For a variation of the approach given below, see [Beall, 2005b], which is closely related to the proposal in [Beall *et al.*, 2005].

[129] Non-normal points were first invoked by Kripke [1965] to model Lewis systems weaker than $4 (systems in which Necessitation fails). [Routley *et al.*, 1982] and [Routley and Loparic, 1978] invoke such points for purposes closer to the current project, as does [Mares, 2004].

from pairs of 'propositions' into go(A/A/). Assuming standard $5 conditions on)A?, conditionals are evaluated thus,

[A —~B[= N UNN

where N, comprising the *normal* worlds at which A ~ B is true, is such that

N)A? if IAI c Igl, and otherwise N 0

and *NN*, the *non-normal* worlds at which A --~ B is true, is given by ¢,

NN = 0(IAI, IBI)

The idea, put (perhaps) more simply, is just this: in addition to 'normal worlds' (among which is the actual), we also recognize *nort-nor'mal wor'lds.* Non-normal worlds are relevant only to the (non-extensional) conditional; extensional connectives behave normally. Conditionals are evaluated differently indeed, entirely arbitrarily at non-normal worlds. In effect, conditionals are evaluated exactly as you would expect at 'normal worlds', but evaluated in any manner one likes at non-normal worlds (provided the given extensional clauses are respected). This would wreak havoc if *validity* were defined over *all* worlds (of all interpretations), but it needn't be, and in fact is not.

Validity is defined in terms of *actual verification* (See §4.2. Note that one could and standardly does say that validity is 'truth-preservation' over all normal worlds, but restricting to just @ makes no difference.) In the present context: A is *actually verified* in an interpretation iff @ c IAI.) So long as there's no interpretation such that ~ E IAI but (e ~ IBI, then A I~ B. Similarly, valid sentences are those that are actually verified on all interpretations. 13°

This simple framework might not yield everything that one wants from a conditional, but, for present purposes, it yields the target desiderata. For example, the conditional is 'genuine', since no interpretation actually verifies A ~ B and A without thereby actually verifying B. Moreover, A --~ A is valid (as brief reflection indicates), and hence all dT-biconditionals are verified. Furthermore, the troubling contraction principles fail, thanks to non-normal worlds. For example, just consider an interpretation according to which @ c IAI and IAI c IBI, but such that w c ~0(IAI,IBI) but w ~ IBI. This serves to invalidate A A (A ~ B) ~ B, and similar counterexamples invalidate the other Curry-generating principles.

There are other approaches that one might take to conditionals in a paraconsistent context, including, of course, tweaking Field's 'neighborhood' approach. For now, the point is simply that there are suitable conditionals available to the dialetheist.

13°Of course, there are broader notions of validity that can be defined, but the given one is most relevant here.

5.2 *Gaps and gluts?*

Priest's dialetheie proposal offers a natural and very simple approach towards dtruth and, in particular, NTP and ECP. On the other hand, one might be open to such 'gluts' but none the less think that there are also *gaps,* sentences that, somehow, are dtruly classified as 'neither dtrue nor dfalse'.[131] Of course, as in §4.1, it makes little sense, and at any rate is incorrect, to say that some sentences are 'neither dtrue nor dfalse', at least given (only) a Strong Kleene account of negation. A natural response to the problem not radically different h'om the *strong truth* tradition is to acknowledge another negation, a more robust (but logically weaker) *gap-closing* negation, in particular, a sort of 'exclusion negation'. The reason that this is not normally pursued is that such a negation one for which LEM holds inevitably gives rise to inconsistency. But this is no major worry if the overall framework is paraconsistent: we may accept LEM for one of the negations (viz., exclusion) but reject EFQ for all negations.[132]

The idea can be modeled using a four-valued language along the lines of Anderson and Belnap's *FDE* [1975; 1992].[133] Our semantic values 12 {1, b, n, 0} are ordered thus:

b:i[n

'0"

Intuitively, as in Priest's proposal, 1 models sentences that are dtrue but not dfalse, 0 sentences that are dfalse but not dtrue, b sentences that are both dtrue and dfalse, and n sentences that are neither. The designated values :D are 1 and b, the idea being, just as in Priest's proposal, that dtrue sentences are designated (even when they are also dfalse).

Interpretations are functions ~ from sentences into F such that L,(A A B) and u(A V B) are the infimum (glb) and supremum (lub) of ~(A) and L,(B), respectively) 34

In *FDE* we have only (what I shall call) 'choice negation' 7, which toggles 1 and 0 and is fixed at both b and n. We add another negation, pseudo-exclusion -, which toggles 1 and 0, is fixed at b, but takes n (gaps) to 1. The result is (what I shall call) FDE*. la5 Accordingly, FDE*-interpretations ~obey' the following diagrams with respect to negation.

---

[131] The material in this section is largely from [Beall, 2005b].

[132] As indicated, paracomplete and paraconsistent theorists unite in rejecting any 'absolute exclusion device', a device Q such that QA, A IF-B and IF-QA V A. The current suggestion is no different.

[133] The name *'FDE'* is now common for the following framework; however, it is perhaps unfortunately so named, since there are various accounts of 'first degree entailment'. But I shall follow what now seems to be common practice.

[134] For present purposes I lay out the propositional semantics; the predicate extension including the resulting dtruth-theory is straightforward: one simply allows for both T + N T fi (3 and T + U T- # S. (See §5.1.)

A		A
0	1	0
n	**1**	**n**
b	b	b
1	**0**	**1**

Notice that dfalsity, following standard thinking, remains dtruth of negation dtruth of *choice* negation, as opposed to pseudo-exclusion (henceforth, exclusion).

A model of A is an FDE*-interpretation that designates A, that is, an interpretation ~ such that ~(A) E ~D. And a model of F = *{As,..., A~}* is a model of Ai, for each 1 ~< i ~< n. Consequence I~ is defined thus: F I~ A iff every model of F is a model of A. Valid sentences are consequences of ~.

With respect to Liars, there will now" be two sorts: choice and exclusion. One could treat both sorts of Liar the same, namely, as gluts. On the other hand, one might follow the methodological principle according to which a sentence is gappy if its truth is determined by neither the world (as it were) nor the language. On such a route, choice Liars similarly, dtruth-tellers (e.g., 'this sentence is dtrue') are treated as gaps, exclusion Liars as gluts. For more on the philosophical picture, see [Beall, 2005b].

5.3 Comments

It is obvious that the foregoing approach has all the virtues of Priest's dialetheism, at least with respect to NTP and ECP (e.g., unified semantically significant predicates, etc.). Moreover, the conditional indicated in §5.1 may be used for purposes of a suitable conditional one that detaches and validates the dT-biconditionals. Here, I touch on two issues. Further discussion is available in [Beall, 2005b].

De Morgan and the notion of gaps

As expected, excluded middle fails for choice negation but holds for exclusion: ~z A V ~A but I~ A V A. Moreover, both negations exhibit standard double-negation behaviour, at least in terms of 'inferences'. For example: A q~ ~A and A q~ A. 136

Standard de Morgan laws hold for choice: ~(A V B) is equivalent to ~A A ~B (and similarly for the other laws). But exclusion is different: de Morgan laws will

---

laSThis is not the best name, as it might suggest an approach to FDE using the Routley star, but I trust that no confusion will ensue.

la6Note, however, that in the double-exclusion case, this is only bi-consequence, not equivalence in the strong sense of 'same value' (which does hold in the choice case). What we have in the exclusion case is co-designation: A and A are both designated or both undesignated on any *FDE*-interpretation.*

generally hold in one direction but not both. Of particular importance given the role of exclusion in the notion of *gaps* is that we have

-(A V B) IP -A A -B

but we do not have equivalence; in fact,

AA (AvS)

A counterexample: p(A) n and t,(B) b. hi that case, p(A) 1 and ~,(B) b, andso,(AA B) b. But, then, p(AVB) i, and so , ((A V B)) 0. ~37

Is the 'nonstandard' behavior of exclusion failure of de Morgan principles a prima facie problem? I see no reason to think as much, in general. Presumably, choice is our 'default' negation; we employ exclusion when we need to talk about failures of choice. Our 'intuitions' about de Morgan, in turn, are presumably based on choice or, at least, based on 'normal cases', 'settled cases', and so on. That some such (de Morgan) principles should fail for exclusion seems, as said, not to be a problem, in general.

On the other hand, one might worry that the given de Morgan 'failures' pose a problem for the role of exclusion in the notion of gaps. Gappy sentences are supposed to be neither dtrue nor dfalse. But that, one would think, ought to be *equivalent* to saying that such sentences are (exclusion-) not dtrue and (exclusion-) not dfalse. The worry is that such equivalence fails, given that, as above, (AVB) and A A B aren't equivalent, in general.

Fortunately, the worry isn't serious: (A V B) and *A A B are* equivalent in the special case where B is ~A, which is precisely the case involved in saying that A is neither dtrue nor dfalse. Accordingly, the general failure of de Morgan (for exclusion) seems not to be a particular problem for the notion of gaps.

Strong truth

If, for some reason, one is inclined to acknowledge a notion of 'strong truth', such a notion is definable in terms of our two negations and dtruth [Beall, 2002]. In particular, define a 'robustly true' device T thus: TA iff dT(~ A}, or, equivalently (given dtruth), TA iff ~ A. ias

6 DTRUTH, VALIDITY, AND TRUTH-PRESERVATION

The notion of validity is often cashed out, at least intuitively, as 'necessary truth-preservation'. At the very least, 'truth preservation' is commonly thought to be

---

137The given 'inference' holds if our values are linearly ordered thus: i ;~ b >~n > 0. But in that case, de Morgan will break down for choice.

laSCompare Field's approach (§4.2).

a necessary condition of validity. But in the context of *dtruth,* the common connection is strained, at best. In this section, I briefly discuss a (relatively under-discussed) issue confronting the two common approaches to dtruth paracomplete and paraconsistent.

One can see the issue in light of the classical picture, wherein we can retain the familiar connection between validity and 'necessary truth-preservation' by giving up a dtruth predicate (for the language) in which to express (in the language) such a connection. (One must resort to a 'richer metalanguage' and, hence, use something other than dtruth.) By contrast, as the foregoing sections have indicated, paracomplete and paraconsistent approaches give us a dtruth predicate; however, the familiar connection between validity and dtruth-preservation is lost (or, at least, strained). I will briefly elaborate. In what follows, I will assume that our basic (extensional) framework is either/(3 (paracomplete) or *LP* (paraconsistent).

6.1 Validity

Validity, as above, is often thought of as *necessary truth-preservation,* or at the very least *requiring* such truth-preservation. A natural way to understand such a thought is as follows, where Val is a binary predicate having the intuitive sense of *follows from.*

V0. Val((d), (B)) ~ s(T(A) ~ T(B))

where --+ is a genuine conditional, one that at least detaches (in 'rule form').[139] In the case of *dtrnth,* which is currently the focus, V0 amounts to

V1. VaI((A), (B)) ~ D(A ~ B)

Intuitive as V1 may be, problems arise in both paracomplete and paraconsistent settings, at least if V1 is supposed to be expressible in one's given language. One problem, arising from Curry, arises for both approaches.

6.2 Curry

To simplify matters, concentrate just on the weaker principle:

V2. VaI((A), (B)) ~ (A ~ B)

As above, we're assuming that ~ is detachable; hence, given plausible assumptions about conjunction, we have

(1) Val((A A (A ~ B)),(B))

---

[139] In the present context, speaking of the 'rule form' of Modus Ponens (or any other argument-form), where this invokes a turnstile, is slightly delicate. After all, a turnstile is typically used in a metalanguage (for a formal object language) to represent what, intuitively, we take to be our validity predicate, which is precisely the issue viz., how to understand a validity predicate *for* our language *in* our language.

But, then, by V2 and detachment we get

(2) *AA(A~ B)~ B*

The trouble is that (2) is a notoriously easy recipe for Curry, which results in triviality. The upshot is that, 'intuitively right' as they may appear, V2 (and, hence, V1) must go.

One not altogether unattractive response to this problem is to reject the dorainant idea that validity requires 'necessary dtruth-preservation' or, more simply, even dtruth-preservation, and indeed reject that 'validity' be defined at all. Perhaps, instead, one ought to accept that *validity* is a *primitive* notion. And perhaps not all is lost: one might be able to retain various 'intuitive' thoughts about validity but reformulate them in terms of 'acceptance' and 'rejection', notions that are already essential to paracomplete and paraconsistent frameworks.[14°]

On the other hand, it would be nicer (in some sense) to retain the idea of 'truth-preservation' if possible. I will consider the paraconsistent and paracomplete cases in turn.

6.3 Paraconsistent: dtruth-preservation

V2 and, hence, V1 must go, as per §6.2. The question is: how, if at all, can we retain the idea that validity requires (let alone *is)* dtruth-preservation?

One route, at least in the *LP* setting, is to give a 'mixed' account of truth-preservation, one that uses some other 'conditional' in addition to one's genuine one. A natural go invokes D, the material conditional. For simplicity, we may concentrate on the simpler case:

V2*. Va,((A), (B)) ~ (A D B)

In an LP-setting, V2* will be dtrue assuming, as I am, various natural ways of handling ~ (e.g., §5.1). Moreover, Curry paradox is harnessed. We have it that detaches, and so have it that (1) is dtrue. But from V2*, unlike the case of V2, we do not get (2); we get only

(3) *AA(A~B) DB*

And since D does not detach in the LP-setting (and, so, isn't a 'genuine conditional'), the threat of triviality is avoided.[141]

[14°] Example: *One ought to reject the validity of a given argument if one accepts that it's possible for the given premises to be true but conclusion false.* This won't generalize to V1 or even V2, at least in the paracomplete framework. (Speaking only of dfalsity will also not quite be enough in the paracomplete case, but I'm simplifying.) Still, in particular cases, such formulations may work. Field (in conversation) has expressed sympathy (if not full endorsement) of this approach.

[141] Let A be a Curry-sentence of the form A ---*±. Since the dT-schema is underwritten with ---*(our 'genuine conditional'), (3) will yield ~A V±, which, in *LP*, is dtrue exactly if A is dfalse, that is, exactly if A ~ ± is dfalse. Provided that sentences of the form A ~ B may be dfalse (at a point) without A's being dtrue (at that point), the threat of Curry-generated triviality is avoided.

One salient cost of retaining dtruth-preservation via V2* concerns the familiar relation between inference (or, indeed, deductions) and validity.[14s] The utility of (knowing that you have) a valid argument is often supposed to be that it is (necessarily) dtruth-preserving. The utility of such dtruth-preservation, in turn, is often supposed to be that our inferences, if dtruth-preserving, 'won't go wrong'. But while V2* ensures that valid arguments are dtruth-preserving (indeed, even necessarily so in *LP)*, the utility of such dtruth-preservation is now dubious: one can know that an argument is valid, that its premises are all dtrue, but none the less be without a valid argument that takes one from such information and the (V2*-) dtruth-preservation of the argument to the given conclusion. In short, the dtruth-preservation of valid arguments is now without any of its familiar pragmatic utility.

To what extent such a cost is crippling remains to be seen. There are (obviously) other ways of 'mixing conditionals' to preserve the idea that validity requires (necessary) dtruth-preservation, but, at least in LP-like settings, most of the options will exhibit the above 'break from pragmatic import'.[14s]

6.4 Paracornplete: (robust) truth-preservation

Just as in the paraconsistent setting, the paracomplete account must give up V2 (and, in turn, V1). The question remains: is there any way to preserve that validity requires dtruth-preservation in a paracomplete setting? The answer is No.

The most promising route is something along the lines of V2*, but that will not work. In *LP*, LEM holds; in K~, not. Without LEM, V2*, regardless of its other costs, is unavailable.[144] Consider, for example, an instance of V2* in which A is the conjunction of a Liar and its negation, and B is '1 = 0'. This immediately shows the inadequacy of V2* in a /{3 setting (wherein EFQ is supposed to be valid).[145]

The upshot is that paracomplete theorists, unlike paraconsistent theorists, cannot enjoy a 'validity' predicate (in the language, for the language) that requires dtruth-preservation. How, then, if at all, can a paracomplete theorist enjoy a 'validity' predicate (for her language, in her language) that, in some sense, is *truth-preserving* (versus dtruth)?

---

[142] Another potential cost concerns the relation between 'counterexamples' (intuitively understood) and validity, but I will ignore this here.

[143] I think that the awkwardness of the given problem can be diminished by drawing a clear distinction between validity (or implication) and inference, much along the lines of Carroll's Tortoise or, more recently, Gil Harman's work [1986]. But I leave the matter there, turning to the paracomplete case.

[144] Of course, since I'm concentrating entirely on *LP*- and Ks-based approaches, my claims need to be taken with a grain of salt. It may be that some paracomplete approaches can utilize something along the lines of V2*. (I doubt it, but I leave it open.)

[145] In the case of the corresponding VI*, the situation is modeled by a case in which v(A • B) 1 ff (undesignated). Taking an obvious route towards the box, wherein p(13A, w) = min{~(A, w~) : *Rww~}*, the same problem is plain.

The obvious idea is to invoke 'robust truth' or 'strong truth' or the like. After all, the purported utility of 'strong truth' is precisely to comment on the status of claims for which the dtruth predicate is inadequate. On this route, we have something like

$$W,((A), (B)) \sim (TAD\ B)$$

I am not clear whether such an approach will work in standard paracomplete frameworks (e.g., Field), but at least in the (admittedly nonstandard) 'paranormal' framework, the approach should work. Whether such an approach will ultimately work, or deliver all that one needs let alone desires with respect to 'truth-preservation', is something I here leave open.[146]

7 LEAVING DTRUTH: PARAMETERS AND CONTEXTUAL SHIFTS

Until now, I have focused entirely on dtruth and the two main approaches towards resolving dtruth-theoretic paradox. I should note that many (if not most) of the well-known approaches towards truth are not accounts of dtruth: they give up the unrestricted intersubstitutivity that is essential to dtruth. One of the main reasons for largely focusing on dtruth is that, as mentioned in §2, it appears to be the toughest case. If one can resolve the dtruth-theoretic paradoxes in an adequate fashion, then, presumably, one can resolve the truth-theoretic paradoxes for other notions of truth, since the latter notions will be less demanding e.g., with respect to unrestricted intersubstitutivity oi" so on. Another reason for focusing on dtruth is that, as mentioned in §1, the other approaches revision theory and what I here call *parametric* approaches are well-covered, or at least clearly sketched, in a variety of accessible places. (See §11.) None the less, something should be said about the other approaches, especially since they are both popular and philosophically (as well as logically) interesting.

The most popular (semantic) approaches to truth at least having left dtruth divide into two very broad camps: *parametric* and *revision.* The former camp contains those accounts according to which NTP is answered via some parameter (hidden or otherwise) associated with truth. The latter camp contains those accounts that answer NTP by positing a 'hypothetical' or 'revision rule' character to truth itself. Parametric accounts of truth, at least in broad outline, are perhaps the best known among philosophers. Revision theory is less known among philosophers, despite being an approach that may well afford numerous philosophical applications.

---

[146] The worry, with respect to Field's framework, is that the proposal (above) calls for a 'unified strong truth' device, something that Field does not have. It *may* be that, in the framework of [Field, 2005b], one could define validity as there being a *good c~* such that, for all /3 > c~, the given argument preserves D/3-truth. But I don't know whether this will work. In the 'paranormal' (similarly 'residual') account (§4.3), we have a unified 'strong truth' device that ought to do the trick.

My aim in the following ~parametric' sections is entirely informah I very (very) briefly sketch the ideas of various sub-camps within the broad, parametric camp. For space reasons, I skip over any formal details, leaving such detail to work cited in §11. Similarly, with respect to revision theory, my aim is merely to give a brief, informal sketch of the idea. Fortunately, there are plenty of works that sufficiently discuss both the informal and formal features of the canvassed accounts. (See references throughout, and §11.)

Alfred Tarski [1956] ushered in not only a renewed exploration of truth, but also the broad, and ongoing, parametric tradition towards truth. As such, I quickly sketch Tarski (or, at least, Tarski*), and then turn to other parametric approaches.

7.1 Tarski: language-parameter

As is well known, Tarski [1956] proposed a hierarchical account of *truth for formal languages.* A language is *semantically closed* just when, for all sentences A in the language, there is a name (A} of A in the language and all T-biconditionals hold: T(A} ~ A.[147] *Tarski's Theorem* is that no consistent, first-order language, which is 'syntactically resourceful' (in effect, can sufficiently 'describe' its own syntax), can be semantically closed, at least where ~ is material equivalence. The reason arises from GSdel's *diagonal lemma,* which applies to such languages: for any open sentence *A(x)* with x free, there is a (true) biconditional B +-~ A((B}), and so Liar-like biconditionals L ~ ~T(L}.

Let £ be a first-order, sufficiently syntactically resourceful, language. How can we have a truth predicate for 12 such that all T-biconditionals hold and 12 is consistent (and, hence, nontrivial)? Tarski's well-known answer invoked a 'richer metalanguage'. In short: do not let T into 12! In other words: forget about chasing after semantically closed languages (theories) ! The result is that Liar-like sentences never arise, and so truth-theoretic paradox (and, in general, semantic-theoretic) paradox is avoided.

Of course, it is one thing to simply 'ban' Liars from a suitably formal language; it is quite another to so ~ban' from a natural language. Tarski's own views on how, if at all, his proposal might apply to natural languages remain unclear.[148] But let us consider Tarski*, a (fictional) character who proposed a Tarskian, hierarchical account of truth for natural languages.

The proposal, in short, runs as follows. We have some (interpreted) semantic-free ~base language' 120- No truth-theoretic (or, in general, semantic-theoretic) paradoxes arise in 120, since the language is devoid of semantic predicates. Let A, B, and C be sentences of £0, and suppose that A and B are true but C false. Of course, since 120 is devoid of a truth predicate, we can't *ezplicitly assert in* 120 that such sentences are true or false. This is the role of the 'next language up', a richer language £1 that contains a truth predicate T1 fox"£0- (Here, I use subscripts and superscripts, respectively, to indicate the languages *.for* and *in* which the predicate is introduced.) If the sentences of 120 are (properly) included among those of 121, one can consistently have it *in* 121 that T~{A} +~ A. (Otherwise, one uses a 'translation' r, where *r(A)* is the translation of A into an £1 sentence, and so in £1 gets the same effect: TI{A} ~ *r(A).)* And this gives us what we wanted: we can explicitly state in £1 the status of our £0 sentences, fox"example, TI{A}, T~{B}, T~{~C}, and more.

---

[147] This is more accurately put in terms of (the technical, logical sense of) *theories,* but I aim here only to give a sketch. The details are available in many places. See §11 for a few sources.

[14S] Tarski seemed to think that truth itself is 'inconsistent' in some fashion, but it is not exactly clear on what he intended. (I have always thought of Tarski as a budding paraconsistentist, but the relevant textual support is weak, to say the least.) For recent discussion of Tarski's views, concerned (unlike here) with precise exegesis, see [Patterson, 2005] and [Soames, 1999] and references therein.

But now what about tile status of such 121 pronouncements? We cannot explicitly assert in 121 that such pronouncements are true; £1 contains no truth predicate for itself. This, in turn, is the role of £1% metalanguage £2, wherein we introduce a predicate T~. And so on. In general, a truth predicate T{~ 1 of level n applies only to the lower-level sentences, sentences of 12.~ for rn < n 1.

Tarski*'s proposal, then, is one that answers NTP by positing a parameter associated with truth: we do not have truth *simpliciter,* but rather only *truth in language such and so.* Tarski*'s parameter, like that of (the real) Tarski, is a language. Nontriviality indeed, consistency is achieved by regimenting out Liars.

What about Tarski*'s answer to ECP? In short: Tarski* gives no answer because he rejects the project. Since, according to Tarski*, we do not have semantically significant predicates *for 12 in 12,* we accordingly do not have 'exhaustive characterization' *in 12,* for any relevant (sufficiently resourceful) language 12. On the other hand, Tarski* certainly does achieve one sense of 'exhaustive characterization', namely, that any suitable *metalanguage* fox"£ can exhaustively characterize £, and do so consistently. Still, it is important to note that Tarski* rejects ECP~ at least in its target sense.

There are various problems associated with Tarski*'s proposal, many of which are well-known. (See §110 Perhaps the biggest problem is the expressive difficulties associated with the proposal. One wants to be able to (competently) use a truth predicate fox"just the sorts of broad generalizations fox"which dtruth was introduced, but in the Tarski* case, things quickly become difficult. For example, in order to (competently, truly) assert that all of So-and-so's claims are truest, one would need to first know that all of So-and-so's claims never reached level n. One could, of course, guess and pick some reasonably high n, but this itself is very difficult, especially if So-and-so is wont to make broad theoretical generalizations concerning language or the like.

A related (and much-discussed) difficultly, made explicit by Kripke [1975], is that some perfectly meaningful or, at least, seemingly perfectly meaningful claims wind up being meaningless on Tarski*'s account. Example: suppose that Nixon says (only) *that everything Dean says is .false,* and similarly that Dean says (only) *that everything Nizon says is true.* Finding suitable levels according to which such claims are meaningful is difficult on a (consistent) Tarskian* approach, to say the least.

There are other problems (both technical and philosophical) with the Tarskian* account, but enough has been said to move on to another, and very popular, approach contextualism.

7.2 Contextual truth

So-called contextualist approaches are increasingly dominant today, at least among philosophically inclined logicians and philosophers of language who are content to depart from dtruth.[149] Part of the attraction may be a tied to currently popular contextualist approaches to vagueness. Whatever the reason, the broad contextualist approach to truth is both popular and, at least in basics, fairly straightforward. Because the contextualist approaches are very (very) broad, space considerations allow only a (very) brief, informal sketch of a few leading ideas. The cited works should be consulted both for fro%her, informal remarks and, in particular, the formal account(s).

In short, contextualist approaches maintain that an interpretation of 'true' depends on context, and does so in a way beyond ordinary ambiguity. (Obviously, if A is ambiguous requiring context to disambiguate -then an ascription of truth to A will likewise be ambiguous. The context-dependence involved in ordinary ambiguity is not at issue.) In particular, Liars seem to generate inconsistency only when we ignore the implicit contextual parameters associated with truth-ascriptions.

Exactly what sort of context-dependence is relevant to paradox is what distinguishes contextualist theories. In its two most popular forms, such context-dependence is cashed out either with an explicit parameter associated with truth or an implicit parameter essentially involved in (at least Liar-like) truth-ascriptions. I will briefly and very broadly sketch these two versions.[15°]

Regardless of the version, contextualists one and all place a lot of weight on the idea of 'reflecting on paradoxicality' and, in turn, 'moving beyond' such paradoxicality. Consider, for example, the following sentences.

$1. $1 is not true.

$2. $1 is not true.

Contextualists find the following reasoning plausible: $1 is true iff not. Hence (given LEM!), $1 is not true, which is what $2 says, in which case $2 is true.[151]

---

[149] Actually, there is no reason a contextualist could not acknowledge a dtruth predicate and yet also proceed to acknowledge other (non-transparent) notions of truth ones not definable in terms of dtruth (and other logical devices). I do not know whether anyone holds such a position, but I think it to be unlikely, since most contextualists are fairly wedded to classical logic. The 'aletheic pluralism' of Crispin Wright [1987; 1992], similarly Michael Lynch [1998], or even that of Michael Dummett [1978], might be a setting for such a position.

[15°] For space reasons, I am omitting discussion of other proposals commonly counted as contextualist e.g., Gaifman, Skyrms, and Koons. See §11.

[151] Strictly speaking, LEM itself can be rejected, but the background logic needs to be 'sufficiently classical'. Contextualists are generally wedded to classical or very-close-to-classical logic.

Notice that the situation is related to 'characterization' and gaps. Consider, for example, the doubly starred sentence:

** The doubly starred sentence is not true.

Assuming the T-biconditionals, which is a common desideratum among contextualists, the doubly starred sentence is true iff not. Hence, given LEM (also a common desideratum), the doubly starred sentence is not true.

But how can such reasoning be sound? In effect, that is the question or, at least, a chief question towards which contextualist proposals are directed, and the question on which I'll briefly focus in §7.2 §7.2.

The foregoing examples are closely related to what Bas van Fraassen [1968; 1970] called *the strengthened Liar* phenomenon. For those call them *caricature gappists* who wish to both reject LEM *and* assert that some sentences are 'neither true not" false' (i.e., that there are gaps), sentences such as the doubly starred sentence serve as strengthened Liars. The caricature gappist proposes that Liars are gappy, and hence not true. But *that*, by all initial appearances, is precisely what the doubly starred sentence says namely, that it is not true. And so it is difficult to see how the proposal achieves both consistency and significant characterization (in the sense at issue throughout). 152

The ~strengthened' phenomenon is not peculiar to caricature gappists; it is quite general, and contextualism, which is in all standard treatments intimately tied to classical logic, is often (if not always) motivated by it. For example, suppose that we introduce a semantically significant (unified) predicate Z for purposes of classifying Liars, and that we impose an *ezclusive* constraint on our semantically significant categories (i.e., that none of the categories overlap). The proposal, then, is that Liars are (in the extension of) Z. A ~strengthened' Liar immediately emerges:

SL. SL is either Z or not true.

Given the appropriate T-sentence and the assumptions above, one soon arrives at the claim that SL is not true.

The question, as above, is how" to make sense of truly and consistently asserting that SL is not true, or that $1 is not true, or that the doubly starred sentence is not true, or so on. How" can any of this make sense in a (more or less) classical framework in which all T-biconditionals hold? *That*, in effect, is the question.

I argue in my pro-hierarchies paper that there is nothing reall to Simmons' non-hierarchical account that is really not hierarchical. It is in a footnote, but the paper is "Truth, Reflection, and Hierarchies", the goal is to show that hierarchies are non-threatening, natural, and everyone should want them, including Keith.

152Ofcourse, at least with *dtruth,* caricature gappism (as it were) makes no sense unless, perhaps as in §5.2, the view is dialetheic (and LEM fails only for one negation or etc.).

Indexical truth

One contextualist tradition stands closely with Tarski*. Tyler Burge [1979] and, in much the same spirit, Keith Simmons [1993] maintain that truth is essentially *indexical.*[153] Truth (or, at least, 'is true'), on this story, behaves indexically: the extension of 'true' is determined by context, so that, in effect, we do not have truth simpliciter, but rather *true in context c.*~54

The idea, in short, is that truth (or 'true') comes equipped with a place-holder or index c for context. In effect, the index is a level in a Tarskian hierarchy, where the given level is determined by context. One context might determine level n, in which case the n-level T-biconditionals are to be used, and in another, level *m*, whereat the m-level T-biconditionals are in play.[155]

How does any of this answer the questions concerning, for example, S1 and \$27 The answer, of course, invokes a change in context, and hence a change in the appropriate indices. In particular, the suggestion is that the (pragmatic) *implicatures* associated in an assertion of \$1 are different from those associated with \$2. When \$1 is asserted in (say) context c, pragmatic principles require that it be evaluated via the c-level T-biconditional:

S1 is truer iff S1 is not truec

But this quickly leads to inconsistency, and hence (more pragmatics) \$1 is said to lack c-level truth conditions, in which case (note well) S1 is not truec. But since \$1 lacks c-level truth conditions, there are no grounds on which to infer, in turn, that \$1 is tru%. Recording as much requires 'reflection' in a different context k /~ c, wherein \$2, which invokes the k-level T-biconditional, may be correctly asserted.

The indexical version of contextualism (be it either Burge's or Simmons') answers NTP by positing a hidden indexical parameter in truth. Classical logic needn't be rejected; one need only recognize that truth is parametric. Non-triviality indeed, consistency is achieved in virtue of the contextually relative extensions of 'true'.

[153] Actually, Simmons' theory purports to be entirely non-hierarchical, unlike Burge's. In that respect, there is a significant difference. For space reasons, I will simply sketch the Burgean idea, but one should note that any reference to 'hierarchies' will be rejected on Simmons' account. Another related approach is in [Koons, 1992]. (On the other hand, Michael Glanzberg [2004c] argues that Simmons' apparent claim to achieve a non-hierarchical theory is both unfounded and unmotivated; he argues that hierarchies are non-threatening, and that Simmons' position is one that should and, in fact, does embrace hierarchies. But this is beyond the current essay.)

[154] Of course, one might try to define some notion of *truth simpliciter* by quantifying over all contexts: A is true simpliciter iff A is true~ for *all* contexts c. But this is generally not countenanced, since (unrestricted) quantification over all contexts makes for inconsistency, at least given classical (or classical enough) logic e.g., 'this sentence is false in all contexts' or the like.

[155] This sketch is potentially more misleading than its mere brevity might otherwise afford, because Burge's overall proposal is a pair of proposals, one a formal theory, the other a set of (very complex) pragmatic principles (that are used to interpret the formal theory). For space reasons, I can only wave at the broad picture, leaving the two essential components (formal and pragmatic) in the background.

What about ECP9 One perennial worry concerns whether such 'indexicalism' achieves exhaustive characterization in the target sense. If the target sense of 'exhaustive characterization' requires a *unified* predicate in terms of which all Liars are properly classified (in the given language), then the Burge Simmons approach seems to fail. Without going into details, suppose that we have a predicate or the same effect, for example via quantification that applies to all Liars. (Intuitively, one might think along the lines of *not true in arty eontezt.)* Let Z be the predicate, and consider SL from §7.2. There seems to be no consistent way of treating such a sentence in the indexicalist's framework. At the very least, some explanation must be given as to why we cannot have such a unified characterization, an explanation that goes beyond pointing to the otherwise resulting inconsistency, an explanation that makes plain why, *despite its apparent intelligibility*, such a notion is incoherent.

Another problem is similar to Tarski*'s (pragmatic) problem. We want to use 'true' for purposes of generalization, but stratification of contexts (on which truth depends) makes this cumbersome, as the pragmatic framework in [Burge, 1979], and similarly (but less explicitly so) in [Simmons, 1993], is quite involved. Indeed, competent, effortless generalizations involving 'true' ultimately require an enormous amount of 'pragmatic know-how', at least on Burge's account. But I will (for space reasons) leave the matter there. 15s

Conteztually sensitive quantifiers

The work of Charles Parsons [1974a; 1974b], who is generally credited with initiating the contemporary contextualist tradition, and Michael Glanzberg [2004a; 2004b], who has clarified and extended Parsons' proposal, represents the other popular contextualist tradition. For convenience, I will call the proposal *quantifier-variability.*

The idea, as before, is that truth ascriptions are essentially context-dependent; however, *there is no indezicality involved in truth (or 'is true').* On the quantifier-variability proposal, the context-dependence the variability involved in truth ascriptions turns on the essential context-dependence of natural language quantitiers. This is an attractive feature of the proposal. Like the indexicalist account, the background logic is assumed to be entirely classical.[157] But importantly unlike the indexicalist, 'true' is entirely univocal, carrying no hidden parameters at all (or ambiguity, or etc.), at least as applied to 'propositions', which, on the current picture, are the 'chief bearers of truth'.

[156] I should note one thing. Burge (similarly, Simmons) takes his account to be *descriptively accurate* of competent usage of 'true' in English. But at least on Burge's account, such usage, as waved at above, involves an enormous amount of pragmatics principles that determine the interpretation of (the level of) 'true'. It is implausible, at least on the surface, that such principles are common knowledge, especially since their relevance arises only with respect to paradoxical claims.

[157] Note that, while I am skipping the formal details here, Glanzberg's work focuses heavily on infinitary languages. While this is essential to some of the particular virtues that Glanzberg claims for his account, it can none the less be set aside for purposes of a broad, basic sketch.

Taking propositions as basic and as the 'things' that are *expressed* by sentences the relevant T-principles have the following form:

If <A} expresses p then p is true iff A

And now consider, for example, the doubly starred sentence in §7.2. The verdict, of course, is that the doubly starred sentence does not express a proposition.

If one sees the basic direction of the proposal (thus far), one will immediately foresee a 'strengthened' problem:

x/~/ The doubly ticked sentence does not express a true proposition.

Given classical logic and the relevant T-principle, the doubly ticked sentence expresses a proposition only if it is both true and not, and so

3. The doubly ticked sentence does not express a proposition.

But, then, familiar (classical) reasoning show's

4. The doubly ticked sentence does not express a true proposition.

And the same familiar (classical) reasoning show's

5. The doubly ticked sentence expresses a true proposition.

What is going on? According to the quantifier-variability proposal, the apparent contradiction is merely apparent, arising from ignoring a contextually shifted quantifier-domain. In particular, there is an implicit existential quantifier at work in the doubly ticked sentence, a quantifier ranging over (a domain of) propositions. The proposal is that both (3) and (5) are true, each being so in virtue of different domains. In the case of (3), the (implicit, existential) quantifier ranges over a domain in which there is no proposition for the doubly ticked sentence to express. In the case of (5), the relevant domain is wider than that of (3), enjoying the requisite proposition.

If one is willing to countenance 'propositions', the quantifier-variability proposal is attractive (at least for those willing to give up dtruth).[158] The story, when filled out (e.g., by Glanzberg), is not only that quantifiers over propositions are context-dependent, but that the background domain of 'truth conditions' (out of which, e.g., 'contents' are often constructed) is similarly subject to contextual expansion such domains get 'bigger and bigger' without end. Accordingly, the 'shifting' at work in the paradoxes is arguably (and so argued by Glanzberg to be) a sort common in linguistic practice, in general, and so not some ad hoc posit invoked merely for paradox.[159]

---

[158] Actually, for those not willing to countenance propositions, Parsons [1983] provides an alternative to the 'proposition' framework, using a truth predicate applying to sentences. (The given predicate may be thought of as being proof-theoretically relativized to a domain of individuals.)

The details, both formal and philosophical, are beyond the current discussion. But the main idea is clear enough. With respect to NTP, the quantifier-variability proposal has it that consistency is achieved in virtue of contextual shifts *not* shifts of an implicitly parametric truth predicate, but rather quantifier-domain shifts. Liar propositions are certainly in many domains of quantification; it's just that they're not in the domains that would otherwise result in inconsistency.

What about ECP9 In short (and very roughly, given the absence of detail), the quantifier-variability proposal rejects ECP, or at least rejects that there is a single context in which we can exhaustively characterize all sentences. The situation is similar in many respects to the indexical version of contextualism (§7.2); the difference is now that there is no 'absolutely unrestricted quantification' with which one would otherwise be able to 'exhaustively characterize' all Liars in the target sense. On the other hand, the quantifier-variability proposal does achieve a great deal. In particular, the proposal seems to be such that for any Liar proposition, there is a context in which the proposition may be truly classified (e.g., as being not true). The price, as mentioned, is the absence of a single context in which to classify all Liars. How steep a price this may be is something that I leave open.[160]

Z3 Situational truth

Jon Barwise and John Etchemendy [1987] propose an account that, in many respects, is similar to the quantifier-variability proposal. Like the latter, truth itself is not indexical in any fashion, and propositions are the chief bearers of truth. Moreover, a sentence (type) may express different propositions depending on context. The difference is that the role of *domains* in the quantifier-variability proposal is now played by *situations,* bits of the world that a statement is about. Hence, propositions are themselves parametric, or at least relative to a given situation. Like the quantifier-variaiblity proposal, 'strengthened' cases for example, S1 and $2 or the like (see §7.2) are resolved by recognizing a contextual shift, one that determines a change in situation and, hence, a corresponding change in the propositions expressed in the given contexts.

The Barwise Etchelnendy proposal which, for convenience, I will call *the situational proposal* is rich in many ways, both philosophically and logically.

---

[159] In Glanzberg's work [2004a; 2004b], the sort of contextual shifts witnessed in Liar reasoning are argued to be common contextual shifts what is 'topical' or 'salient' at a given stage in a discourse and not merely peculiar to truth-theoretic paradox. It is precisely this broader phenomenon of 'shifting' that, according to Glanzberg, induces the mentioned 'bigger and bigger' expansions of the background domain of 'truth conditions'. (As mentioned above, Glanzberg works with infinitary languages and, by doing so, purports to make explicit the content of various Liar phenomena. In the formal framework, Glanzberg develops levels of a hierarchy that are as semantically closed as Kripke's minimal fixed point.)

[160] Another advantage of the quantifier-variability proposal is that it affords a natural and entirely analogous resolution of 'set'-theoretic paradoxes. See [Parsons, 1974b] for discussion. (And see §9 for the reasons behind scare-quoting 'set'.)

A clear account of the proposal is difficult to give without discussing the formal framework, which, in this case, relies on Peter Aczel's ZFC/AFA non-well-founded set theory [1988]. Such a set theory rejects the usual axiom of foundation, and hence allows for 'self-membership' (among other sorts of circularity or, more generally, non-well-foundedness); this allows for (modeling) self-referential propositions in a natural way. For space reasons, I cannot lay out the background set theory or its key theorems, and so cannot go into the formal details of the situational proposal. Instead, I will give a very brief snapshot of the basic idea, leaving cited sources for a fuller and more accurate picture.

As above, the heart of the situational proposal is a contextually sensitive feature of propositions. Two factors combine to determine what proposition is expressed by a sentence on a given use: namely, *demonstrative conventions* and *descriptive conventions*. The former pick out a situation s, and the latter a *type T* of situation. Once these elements are determined, a proposition is at hand, namely, {s;T}, which is the proposition that s is of type T. This account of propositions and their 'making' is what Barwise Etchemendy call the *Austinian* account,[161] where the corresponding (Austinian) account of truth is then natural:

{s; T} is true iff s is of type T

Situations, as above, are bits of the world that a statement is about. To be (slightly) more precise, the *world* consists of *states of affairs*, which, in turn, are built from individuals and properties (or relations); and a set of such states of affairs is a situation. So, in general, states of affairs have the form (R~, 01,..., 0~; i), where R~ is an n-ary relation (a property if n 1) and i is a polarity, intuitively, a truth value out of {1, 0}. If *SA* is the class of states of affairs, then any situation s is a subset of SA. Moreover, states of affairs yield types of situations: if cr E SA, then [G] is a *type of situation*. Hence, a type T looks like this: [(R~, 01,..., 0~; i}]. Let ~ be the state of affairs that determines T [or1. Then the account of truth, above, is equivalent to:

{s; T} is true iff cr c s

We say that s is of (simple) type [~1 exactly if G E s, and so {s; T} is true iff s is of type T.[162]

One might wonder about the role of *the world* in all of this. The role of the world is to distinguish between *accessible* and *inaccessible* propositions. If p {s; T} for some T, then p is *about* s. A proposition is accessible if it is about an *actual* situation, and inaccessible if about a *non-actual* situation. This distinction plays no part in determining the truth (or falsity) of propositions, since there are true (similarly, false) propositions about both actual and non-actual situations. (This is all made abundantly clear in the formal model, but for present purposes, an informal example will suffice.) Simplifying, if (F, y) is a state of affairs that is not part of the actual world, any accessible proposition that claims (as it were) that y is F, is a false (but accessible) proposition, since any such accessible proposition is one about situations that are not of the required type. On the other hand, if (F, y) is a state of affairs that is part of the actual world, then any accessible proposition that claims (as it were) that y is not F, is likewise false, since the proposition is about an actual situation that is of the wrong type.

---

[161] This is in contrast to the *Russellian* account, which can also be modeled in the proposed framework. For space reasons, I will restrict the sketch to Austinian propositions.

[162] Barwise Etchemendy also define molecular types, for example, c, is a disjunctive type [A XI exactly if (7 is of some type T ~ X. Similarly for conjunctive types (dual).

The import for simple Liars, for example, *that this proposition is false*, is that they are simply false if about actual situations. For example, a simple Liar proposition ¢:~ = {s; *Etrue,¢:~;o]}*, if accessible, is false because about a situation for which there are no states of affairs of the right type. On the other hand, given situations that contain (e.g.) *(true, fs;* 1}, some Liars are true; it's just that no such situation is (or can be) actual, on pain of triviality.

As with the quantifier-variability proposal, contextual shifts are invoked to explain the sort of reasoning involved in typical 'strengthened' cases at least, for those, like $1 and $2, that can express propositions. Let fs, as above, be a Liar about actual situation s. Then f~ is false (for reasons above). But how, then, do we so characterize f~? We 'expand' s to s', where s' contains the fact (state of affairs) of f~'s falsity. The proposition that we use to characterize f~ is *p* *{s';* [tr'ue, fs; 0]}, which is that s2 is of f~'s type (where f~ is about s), and hence is true because s2 contains *(true, f~;O}*. But now s' itself has Liars, for example, *f~, = {s'; [tT"ue,fs,;* 0]}. But *f~,* is handled in a similar manner: expand to s", a situation, unlike s', that contains the fact that *f~,* is false. Just like the quantifier-variability proposal, the picture is one in which a sequence of propositions emerges, a sequence featuring alternating truth values that reflect contextual shifts: *¢~, p, f~,, p'*, and so on. We achieve consistency by switching to different (broader) situations in which the 'previous' statements are evaluated.

The situational proposal, then, is one according to which NTP is answered by relativizing propositions to situations. When one pays heed to the situation a Liar is about, the apparent inconsistency generated by Liars is seen to be merely apparent: the given situations are not of the right type. And the proposal shares many (if not all) of the virtues enjoyed by the quantifier-variability proposal, similarly adverting to a phenomenon contextual shift that is arguably a very general one, rather than something peculiar to paradoxes.

With respect to Ecp, the situational proposal faces the same problem (or, at least, same sort of problem) confronted by other contextualist proposals. Indeed, though the details are different, the general issue confronting the quantifier-variability proposal emerges for the situational account: in particular, there is no 'global' situation in which we can truly characterize all Liars. Put another way, the world itself is not a situation about which we can consistently talk. (If we could, then there would be global-Liars, and we would be unable to 'expand' to a bigger situation.) But I leave the matter there.[163]

---

[163] I should note that Barwise Etchemendy address the given worry in a postscript [1987]. They propose a way in which we can talk about the whole world, that is, have propositions about the

8 REVISION THEORY

Like the parametric and contextual approaches, revision theorists leave dtruth behind by giving up on the full intersubstitutivity of T(A} and A. But revision theory is significantly different from the foregoing approaches, at least in its philosophical motivation. To fully appreciate the theory (or family of theories), the technical details are important. For space reasons, I cannot go into the technical machinery; instead, my aim is simply to convey the basic idea.

The two pioneers of revision theory are Hans Herzberger and Anil Gupta, who independently came up with the technical ideas in the early eighties. Herzberger's theory [1982] was motivated by an attempt to 'classicalize' Kripke's inductive approach to truth, the aim being a modification affording classical, two-valued valuations, one according to which Liars (or the like) were neither gaps nor gluts, but rather *unstable.* The given 'instability' is reflected in the manner in which sentences are interpreted. Specifically, sentences are evaluated classically but via a sequence of stages; on some such stages, paradoxical sentences are true, and on others false. One of the chief aims of Herzberger's theory is to characterize just such patterns of instability.

Gupta [1982] and, in turn, Gupta Belnap [1993], motivate a revision theoretic approach to *truth* as an instance of a much broader revision theoretic approach to *definitions,* in general. For present purposes, I will focus on the Gupta Belnap proposal I1993], which provides a much broader philosophical picture than found in Herzberger.[164]

On the traditional account of definitions, a proper definition is neither *circular* not" *creative.* A definition is circular if the definiendum appears in the definiens. A definition is creative if it yields proofs of claims that are not reducible to definiendum-free claims. While revision theory maintains the stricture against creativity, the proposal breaks from tradition by allowing circular definitions. Indeed, in large part, the point of revision theory, at least qua theory of definitions, is to allow" for meaningful, useful circular definitions. Gupta Belnap propose to preserve classical logic, and so we assume as much in the background.[165]

Consider an example that nicely illustrates the general idea [Gupta and Belnap, 1993]. Suppose that we are given the following definition of G, and that our domain is $\{a, b, c, d\}$.

---

world at large. The result is that we can sometimes talk about the whole world, but any Liar-sentence expresses a proposition only about a proper part of the world. On the surface, this is philosophically suspect, since once we are able to talk about the whole world, it is very difficult to see how global-Liars do not emerge. Indeed, once talk of the whole world is available, one could introduce a convention that indicates global talk, some symbol that indicates a global reading of the proposition.

[164] This is not a criticism of Herzberger's work. The point is that, for present purposes, the broader philosophical perspective is appropriate.

[165] It should be noted that revision theory, at least the basic idea, is compatible with many nonclassical logics. Indeed, see the 'revision flavor' of Field's conditional, which is clear in the 'restricted semantics' (§4.2).

Gx =~f (Fx n Hx) V (Fx n ~Hx n Cx) V (~Fx n Hx n ~Gz)

Suppose, too, that we have the following information:

i. a and b are (in the extension of) F.

ii. c and d are not F.

iii. a and c are H.

iv. b and d are not H.

How might we determine the Gs? How, in other words, might we go about discovering the extension of G? The obvious problem confronting our definition of G is that it is circular. Figuring out the extension of G requires, it seems, already knowing the extension of G. So it appears.

Revision theory enters with a simple (but powerful) solution: posit a *hypothetical* extension and proceed to calculate the extension of G on the basis of the hypothesis! Let h0 be your initial such hypothesis. Your subsequent calculation will yield an extension of *G relative to* h0, an extension that, depending on h0, may well differ from h0, and hence *revise* your initial hypothesis.

Consider, for example, an initial hypothesis h0 according to which Q, the extension of G, is 0. With respect to object a, we immediately have it that a c Gh°, that is, that a is in *the extension of G relative to ho,* since, given (i) and (iii), we have the truth of *Fa A Ha,* and so the a-instance of the entire (disjunctive) definiens, which is evaluated classically, is true. Likewise, it is clear that c E Gh°, since, given (ii) and (iii), we have the truth of *~Fc A Hc A ~Gc,* and so the truth of the c-instance of the entire definiens. Similar calculation shows that neither b nor d is in Qh,~. Hence, relative to our initial (in this case, null) hypothesis h0, we have come upon an extension for G, one that is relative to, and distinct from, h0, namely {a, d}.

But what now? After all, h0 was just a blind guess! The next step, according to revision theory, is to start again, but this time running with the 'discovered' extension ~ho as one's hypothesis (i.e., hypothetical extension). So, in the 'second run', hi is the hypothesis that ~ {a, d}. Simple calculation based on hi and (i) (iv) will yield an extension ~ hi , tile extension of *G relative to hypothesis hi.* And so on.

Thinking of n as the *stage* at which h,~ is posited, the general recipe though eventually (but not here) carried into the transfinite is basically as follows. One begins with an initial hypothesis h0 to get an extension $h(, that is relative to (but frequently different from) h0. At stage 1, one's hypothesis hi is that the relevant extension is gh(,. At stage 2, one's hypothesis *h2* is that the relevant extension is £hi. In general, one's hypothesis at stage n + 1 is that the relevant extension is $h., an extension 'discovered' by calculation relative to h,~.16~

166Because I am omitting the formal details, I am also omitting discussion of the revision

The foregoing recipe, the process, generates a so-called *revision sequence*, which is a sequence of ~candidate extensions' for the relevant predicate. Each such revision sequence is *relative to the initial, stage-one hypothesis.* Consider, again, the case of G above. For n > 0, let (n,:c} represent *stage* n at which the extension of G relative to h~ 1 is z. (In other words, z is the extension one gets at stage n + 1 by using the n-stage extension as initial input.) Then the revision sequence, *relative to our" initial (null) hypothesis ho,* looks as follows.

(0, 0), (2, (a, (4,

This sequence, as said, is relative to our given h0, in this case, an initial hypothesis according to which nothing is G. But, of course, with a domain of four objects, there are sixteen possible starting points, sixteen possible initial hypotheses (only one of which is null). Canvassing *all* such resulting sequences exhibits the sort of 'patterns of (in-) stability' that revision theorists seek to illuminate. Two notions are central:

S. Sentence B is *stable* if, after various stages, B forevermore enjoys the same value (either true or false).

C. Sentence B exhibits *convergence* if B stabilizes on the same value in all sequences.

In the current case, given the sixteen possible starting points and the resulting revision sequences based on (i) (iv), the following patterns emerge:~67

R1. Object a. If h0, the initial (not necessarily null) hypothesis has it that a c ~, then a c ~h~ for alln. If h0 has it that a ~ ~, then a c ~ for alln > 0. Accordingly, the sentence *Ga* is eventually forevermore true regardless of starting point, and hence both *stable* and *convergent.*

R2. Object b. If h0 is such that b E G, then b E Gh'~ for all n, in which case *Gb* stabilizes on all such sequences (sharing the ~positive' b-hypothesis). Similarly, if h0 has it that b ~ G, then b ~ Gh" for all n, in which case, again, *Gb* stablizes. So, *Gb* is stable on all initial hypotheses, but, since it doesn't stabilize on the same value in all sequences, it fails to be convergent.

R3. Object c. There is no stability with respect to *Oc,* and hence no convergence.

R4. Object d. This is exactly like the case of a, except that *Gd* is converges on *falsity,* rather than *truth.*

---

process defined over transfinite ordinals. What should be noted is that different revision theories largely differ with respect to their treatment of limit ordinals and, as a result, how 'categorical' judgments about extensions are garnered. See §11 for relevant works.

[167] Such patterns are best conveyed graphically, but for present purposes I merely sketch the import.

With such patterns at hand, and the corresponding features of stability and convergence, revision theory finally arrives at so-called categorical verdicts, or 'absolute values', as one might say. After all, if the revision process is to be of any use, we want to arrive at a *single* judgment about (say) G's extension, as opposed to a bunch of judgments based entirely on arbitrary starting points. Not surprisingly, this is precisely the role of *conver'gence.* In the case at hand, we conclude that a is categorically G and that d is categorically non-G. And those, at least in the given case, exhaust the available categorical judgments. Nothing categorical can be said of c, as it remains simply *unstable.* Similarly, nothing categorical can be said of b, since, despite stability in each sequence, there is no convergence. Convergence gives us *a single value irrespective of starting point.* Failing that, it would be unreasonable to make a categorical judgment. And so it is with revision theory.

What does any of this have to do with *truth?* The answer, in short, is that truth itself is said to be a circular concept. In particular, Gupt~Belnap posit an ambiguity in the usual T-biconditionals. Such biconditionals can be read either *definitionally* or *materially* (i.e., material equivalence). Each such biconditional, when *definitionally* read, has the form

T(A~}   df A~

for i E N. But, then, since A~, for some i, may contain T, the resulting definition of T, namely,

T(~'g')   *df* ('~"   (A0}/~ A0) V (x   /A1}/~ A1) V (95"   (A1}/~ A1) V...

will be circular, the definiendum appearing in the definiens, lss

The basic picture, then, is clear. Our definition of truth which relies on the definitional T-biconditionals is circular. Revision theory is a theory of circular definitions. The proposal is that *the essential character of truth,* its essential definition, is a *rule of revision,* a rule that tells us how to go from 'initial hypotheses' to new hypotheses about the extension of 'true', a rule that ultimately yields useful *categorical judgments* about truth about the extension of 'true'. Just as in the (abstract) example of G, so too with truth: some sentences will converge, others stabilize but not converge, and others never stabilize at all. Paradoxical sentences, at least ones that would otherwise result in inconsistency, fail to stabilize.

The exact judgments concerning Liars and other 'ungrounded' sentences depend on the particular revision theory. (See §11 for relevant works.) But the general proposal, as above, is clear. With respect to NTP, for example, the revision theoretic proposal is that nontriviality indeed, consistency is achieved in virtue of truth's ~revision-rule character', one that results in certain sentences (e.g., Liars) being unstable.

Unlike the traditional account of definitions, according to which circularity often results in inconsistency, appropriate revision rules serve to allow meaningful

---

168I should mention, but (for space) cannot discuss, that Gupta Belnap prove various results about the behavior of d/ (as it were) and -- (material equivalence), showing that they are the same except in abnormal cases. See [Gupta and Belnap, 1993] for full discussion.

and *useful* circular definitions without inconsistency. A nice feature of revision theoretic approaches is that, as in the quantifier-variability and situational approaches, a *general* phenomenon is invoked to explain paradox, rather than some ad hoc feature peculiar to paradox. In the present case, the general phenomenon is circular definitions.

What about ECP? In short, revision theorists reject the project. Any version of revision theory (as far as I can see) will invoke semantically significant categories that, were they introduced into the (object-) language, would result in inconsistency. Consider, for example, a typical 'strengthened' case like 'this sentence is not true or unstable'. This sentence poses problems. Of course, one could, as suggested in [Gupta and Belnap, 1993], introduce ~unstable' into the language and, in turn, have a richer metalanguage that categorized the problem sentence as (say) 'super-unstable' or the like. But, again, nothing close to the (target sense) of exhaustive characterization is achieved. On the other hand, it may well be, as argued in [Gupta and Belnap, 1993], that for any stage s of (say) English, there is a stage sI in which s-English is exhaustively characterized. What one won't have at least not consistently is some stage s such that s-English exhaustively characterizes st-English *for all st.* As with other proposals, I leave the philosophical cost-benefit analysis of the matter open.[169]

9 SET-THEORETIC PARADOXES?

For space reasons, this essay has not covered the wide range of so-called semantic paradoxes that, in many ways, are closely related to truth-theoretic paradox for example, denotation, satisfaction, reference, etc. Instead, the focus has been on the Liar, and a limited number of approaches to the Liar.[170] But something should be said about so-called set-theoretic paradoxes, which, despite Ramsey's division [1925], are often thought to be in kind with the Liar. For present purposes, I will only give a brief comment.

Russell's paradox is often thought to be the set-theoretic counterpart of the Liar. The naïve comprehension scheme, according to which, for any open sentence A(x), there is a set y such that o c 3; iff *A(o)*, for any object o. Russell's paradox invokes z ~ z. By the comprehension scheme, we have some set P~ comprising all and only the objects that are 'non-self-membered', that is, the objects satisfying x ~ -c. But, then, 7~ c 7~ iff 7~ ~ 7~. The resemblance to Liars and the T-schema is obvious.

---

[169] I should make it plain that Gupt~Belnap [1993] argue that ECP, inasmuch as its target sense overlaps with what they call 'semantic self-sufficiency', is unfounded. My own thinking is that we *star't* with the strong appearance of 'exhaustive characterization', that (e.g.) natural languages are exhaustively characterizable in terms of unified (semantically significant) devices. If we cannot make sense of how this could be if, that is, we cannot answer ECP adequately then we should look for alternatives, perhaps along the lines of the 'stages of English' suggestion above (where the order of quantifiers is critical). But, for repeated space reasons, I must leave the matter there.

[170] Indeed, I have not covered even the wider family of truth-theoretic paradoxes, those paradoxes (e.g., Knower, Necessity, etc.) that involve 'truth-entailing notions'.

What should be noted is that, in effect, *Russell's paradox* amounts to two paradoxes, one for *set theory,* one for *semantics.* By my lights, mathematics is free to simply axiomatize away its version of Russell's paradox, as is standardly done. Sets were introduced within and for mathematics, a role that need not be constrained by our 'intuitions' about natural language. So long as the given role is sufficiently served by %xiomatically harnessed' sets, then so be it. Of course, mathematicians frequently seek the most natural or 'beautiful' route in their endeavors, and perhaps standard axiomatized approaches towards sets are not as natural as one nfight like. But that is a separate issue. The main goal is to find entities that play the given nlathelnatical role.

Semantics, on the other hand, seems to be a different matter. What semantics needs is a theory of (what I will call) *semantical properties.*[171] Semantical properties are entities that play a particular semantic role, in particular, those entities corresponding to each meaningful predicate in the (given) language. The role of semantical properties is essentially given by the naive semantical schema: namely, that for any object o and predicate Y(~c), there is a semantical property ~Y such that o exemplifies £Y iff *Y(o)* is true. This principle is fundamental to semantics, at least in practice. But the principle confronts Russell's paradox for *semantics,* in particular, the instance invoking :c *does not ezemplify ~c'.*

Perhaps one reason that the two paradoxes are not distinguished is that semanticists simply borrowed the mathematical sets to play the role of semantical properties, and the history of (at least contemporary, formal) semantics has gone along with the borrowed entities. But why think that the mathematicians' sets will sufficiently play the role of semantical properties? There is no a priori reason to think as much. Indeed, inasmuch as semantics needs its unrestricted naive semantical schema (above), there is reason to be pessimistic.

A solution to the Liar, at least by my lights, need not thereby be a solution to *mathematical set-theoretic* paradoxes; those paradoxes may be solved in whatever way suits mathematics. Of course, if mathematics remains classical, then one's proposed %ll-purpose logic' will need to have classical logic as an extension. The point is that Russell's set-theoretic paradox needn't be an issue in one's resolution of the Liar or truth-theoretic paradoxes, in general. On the other hand, one's resolution of the Liar is constrained, to some degree, by Russell's semantics-theoretic paradox the semantical properties paradox. At the very least, solutions to the Liar that require rejecting the unrestricted semantical schema would seem to carry a prima facie blemish.

As it turns out, accommodating the unrestricted semantical schema usually goes hand in hand with achieving a suitable conditional, one for which A ~ A is valid,

---

[171]i used to use the term 'semantical extensions', but Hartry Field (in conversation) convinced me that 'extensions' is apt to mislead, perhaps suggesting the target entities are *extensional. I* use Field's suggested term 'properties', which is more natural.

--~ detaches, and Curry problems avoided, iT> Accordingly, some of the theories sketched above, notably the paracomplete and paraconsistent, naturally afford a solution to Russell's semantical paradox. 17a

10 REVENGE AT LAST

Charges of *revenge* are ubiquitous in the Liar literature. The loose metaphor aptly conveys the target phenomenon, fuzzy as the phenomenon may be. Dramatically put, Liars attempt to wreak inconsistency (or triviality) in one's language. One way of avoiding the Liar's intentions is to kill the Liars, much along Tarski*'s line or lines (not discussed here) that treat them as meaningless. But killing Liars, at least natural-language Liars, carries costs, as mentioned in §7.1. The more humane approach, exhibited by all of the canvassed approaches (except Tarski*), is to let Liars live but frustrate their aim of generating inconsistency (or triviality). And this is where the Liar's revenge is thought to emerge. In short, the Liar's aims are safely frustrated only at an expressive cost. If the Liar can't have what she wants, she'll enlist ~strengthened' relatives to frustrate your wants, in particular, your expressive wants.

So goes the drama. In less dramatic terms, *revengers* charge that a given (formal) theory of truth fails to be explanatorily adequate with respect to its target phenomena. In all cases, the target phenomena include at least the consistency or, more broadly, nontriviality of truth, *truth* in our *real* language. ~74 The revenger charges that the formal 'model language' achieves consistency (or nontriviality) only in virtue of expressive inadequacies. In particular, some semantic

---

[172] I should note that the semantical properties version of Curry is directly analogous to the truth-theoretic version. Let e be the exemplification relation, [x : A(z)] the semantical property corresponding to *A(z),* and *xe[y* : A(y)] +-~ *A(z)* the semanticM property schema. Curry is generated via the predicate x C z + ±, yielding [z : x C x + ±], where ± is either explosive or plainly false.

[173] For concrete examples, see [Field, 2004] and [Priest, 1987]. I should note that Priest [1987] argues that any adequate solution to the Liar must be a solution to Russell's set-theoretic paradox. I reject the argument, but the matter ultimately turns on whether there is a sustainable distinction between semanticM properties and mathematicM sets. Priest (in conversation) thinks not, but also thinks that mathematics itself is ultimately inconsistent, thereby calling for a paraconsistent logic. These issues, needless to say, are (far) beyond the scope of this essay.

[174] All of this assumes, as I have throughout, that we are dealing with proposals that purport to be *descriptively adequate* with respect to various target phenomena. Those, like McGee [1991], who think that our real language is inconsistent due to Liars and, in turn, proceed to offer a chiefly *prescriptive* (consistent) remedy, are not covered in this discussion. With respect to the current issue, McGee's position is somewhat curious. If, as McGee [1991] concludes (on the descriptive side), English is inconsistent, then why think that its logic is explosive (e.g., classical)? Presumably, it is because the logic is explosive that a prescriptive remedy is required. But a more reasonable hypothesis, at least on the surface, is that the logic is paraconsistent, at least if as McGee maintains our 'ha'i've theory of truth' is inconsistent. But if the logic is paraconsistent, then why do we need a consistent remedy? But I must leave the matter there. (I should mention that McGee's work, which, only for space reasons, I have not been able to cover, is among the pioneering works in the 'determinate truth' tradition.)

notion 2(, which is expressible in our real language, is, *on pain of inconsistency (or triviality),* inexpressible in the (formal) model language.

Doing formal semantics is modeling work, especially when one is engaged in philosophical semantics or philosophical logic, where the aim, in general, is to illuminate the structm'e and logic of philosophically significant notions semantic notions (truth, language, etc.), metaphysical notions (worlds, parts, etc.), perhaps ethical notions, or so on. With respect to formal theories of truth, the aim is no different; one aims to give a model that sutticiently illuminates various features of truth. Of course, in the case of truth, one aims to give a *model* in more than the ordinary, heuristic sense (e.g., exemplar); one also aims to establish *consistency* or, more broadly, *nontriviality.*

In giving a formal theory of truth, one does not directly give a theory of *truth;* rather, one gives a theory of *truth-in-/2,~,* for some formal 'model language'/2,~. By endorsing a formal theory of truth, one is endorsing that *truth* itself is relevantly like *that,* like truth-in-/2,~, at least with respect to various phenomena in question for example, truth's consistency (or nontriviality), its logic, etc. Revengers charge that truth-in-/2~ is insutticiently like truth for purposes of explaining the target phenomena.

For simplicity, let /2,~ be a given formal model language for/2, where/2 is our target, real language the language features of which (e.g., consistent truth) £,~ is intended to explain or otherwise illuminate. Let *M(/2,~)* be the metalanguage for £,~, and assume, as is typical, that *M(/2,~)* is a fragment of/2. Then various (related) recipes for revenge run roughly as follows.[175]

Rvl. RECIPE ONE.

- Find some semantic notion 32 that is used in *M(£,~)* to classify various /2.~-sentences (usually, paradoxical sentences).

  Show, in *M(g~r,),* that 32 is not expressible in/2~r~lest/2~ be inconsistent (or trivial).

  Conclude that £.~ is explanatorily inadequate: it fails to explain how £, with its semantic notion 32, enjoys consistency (or, more broadly, nontriviality).

Rv2. RECIPE TWO.

  Find some semantic notion 2(that, irrespective of whether it is explicitly used to classify £,~-sentences, is expressible in M(£,~).

- Show, in M(/2,~), that 2(is not expressible in £,~ lest £,~ be inconsistent (or trivial).

---

[175] This is not in any way an exhaustive list of recipes! (A related route, for example, is pursued by Glanzberg [2001], who argues that any 'good' theory of truth cannot be *applied* without invoking devices that, were they 'allowed' in the theory, would result in inconsistency or triviality. A proof-theoretic version of this route towards 'revenge' is in [Glanzberg, 2004c].)

Conclude that 12m is explanatorily inadequate: it fails to explain how 12, with its semantic notion X, enjoys consistency (or, more broadly, nontriviality).

Rv3. RECIPE THaEE.

Find some semantic notion X that is (allegedly) in 12.

Argue that 2(is not expressible in 12.~ lest 12,~ be inconsistent (or trivial).

Conclude that 12rr~is explanatorily inadequate: it fails to explain how 12, with its semantic notion X, enjoys consistency (or, more broadly, nontriviality).

A revenger, as above, is one who charges 'revenge' against a formal theory of truth, usually along lines above. The charge is that the model language fails to achieve its explanatory goals. Suppose that 12.~ is some formal language that enjoys its own truth predicate. Suppose, further, that consistency (or, for paraconsistentists, nontriviality) is established for the given language (and the resulting truth theory). A *rever~ger"* charges that 12.~ fails to explain its target explanada, that, in particular, its consistency (or nontriviality) is achieved in virtue of expressive inadequacies. In general, the revenger aims to show that there's some sentence in 12 (our real language) that *ought* to be expressible in 12~r~if 12~ is to achieve explanatory adequacy.

How ought one to reply to revengers? The answer, of course, depends on the details of the given theories and the given charge of revenge. For present purposes, without going into such details, a few general remarks can be made.

First, the weight of (Rvl) or (Rv2) depends on the sort of?(at issue. As in §4.1 §5.1, if X is a model-relative notion constructed in a *proper'fT"agment of 12,* then the charge of inadequacy would seem to carry little weight. In particular, as mentioned in §4.1 §5.1, if classical logic extends that of 12,~, then there is a clear sense in which you may 'properly' rely on a classical metalanguage in constructing 12.~ and, in particular, truth-in-12"~. After all, you endorse that 12, the real, target language, is nonclassical but enjoys classical logic as a (proper) extension. Accordingly, while details need to be given, there is nothing suspect about relying on an entirely classical (proper) fragment of 12 in which to construct your model language.[176] But in such a context, it is hardly surprising that 2(, being an entirely classical notion (at least over 12.~), would bring about inconsistency or, worse, triviality, if we were to express it in £rn.[177]

Because classical logic is typically an extension of the logic of 12m, the point above is often sufficient to undermine the revenger's claims, at least if the given recipe is (Rvl) and (Rv2). What the revenger must show, ultimately, is not the unsurprising result that a model-relative X is expressible in *M(£,,~)* but not in £,~; she must show that some relevant non-model-relative notion is expressible in £ but, on pain of inconsistency (or triviality), inexpressible in Z;~r~. And this task brings us to (Rv3). lr8

---

[176] Note too that if the given fragment is 'necessarily classical' in some relevant sense, then inferences within the given (proper) fragment of £ may turn out to be valid simplieiter depending on how this is cashed out. Otherwise, some restricted but, as far as I can see, entirely appropriate notion of validity will be in play for the 'metalanguage inferences'.

[177] Hartry Field [2003b] discusses this point with respect to his particular proposal. For a fuller discussion, see [Field, 2005b].

Recipe (Rv3) is perhaps what most revengers are following. In this case, the idea is to locate a non-model-relative (semantic) notion in Z; and show that Z;,~ cannot, on pain of inconsistency (or triviality), express such a notion. As may be evident, the dialectic along these lines is delicate. Suppose that Theorist proposes some formal theory of truth, and Revenger, following (Rv3), adverts to some non-model-relative (semantic) X that (allegedly) is expressible in /2. If, as I'm now assuming, Theorist neither explicitly nor implicitly invokes ,12 for purposes of (semantic) classification, then Revenger has a formidable task in front of her. In particular, Revenger must, without begging the question, show that X really is an intelligible notion of £.

An example, relevant to the paracolnplete and paraconsistent cases (see §4 §5), might be useful. Let us say that a language contains an *absohtte ezcl,tsion device* (P exactly if *both* of the following hold for the given language.

edl. IP qPA V A

ed2. qPA, A I~ B

Against a paracomplete or paraconsistent proposal, a revenger, following (Rv3), will not make the mistake of pointing to some model-relative exclusion device. Rather, the (Rv3)-type revenger might maintain that £, our real (and target) language, enjoys such an absolute exclusion device. If the revenger is correct, then standard paracomplete and paraconsistent proposals are inadequate, to say the least. But the issue is: why think that the revenger is correct? Needless to say, argument is required. What makes the matter delicate is that many arguments are likely to beg the question at hand. After all, according to paracomplete and paraconsistent theorists, what the Liar teaches us is that there is no absolute exclusion device. Accordingly, the given revenger cannot simply point to normal evidence for such a device and take that to be sufficient. On the other hand, if the proposed theory cannot otherwise explain or, perhaps, explain away normal evidence for the (alleged) device, then the revenger may make progress. But the situation, as said, is delicate.

The burden, of course, lies not only on Revenger; it also lies with the given Theorist. Paracomplete and paraconsistent theorists must reject the intelligibility of any absolute exclusion device. But inasmuch as such a notion is independently plausible or, at least, independently intelligible such theorists carry the burden of explaining why such a notion appears to be intelligible, despite its ultimate unintelligibility. (E.g., we are making a common, reasonable, but ultimately

---

lrSThere is (obviously) a lot more that can be said about (Rvl) or (Rv2), but for space reasons I briefly turn to (Rv3).

fallacious generalization from ~normal cases' to all cases, or some such mistake.) Alternatively, such theorists might argue that, contrary to initial appearances, the allegedly intelligible notion is in fact rather unclear; once clarified, the problematic features disappear. (E.g., one might argue that the alleged notion is a conflation of various notions, each one of which is intelligible but not one of which behaves in the alleged, problematic way.) Whatever the response, theorists do owe something to (Rv3)-revengers: an explanation as to why the given (and otherwise problematic) notion is unintelligible.[179]

I have hardly scratched the surface of *r'evenge* in the foregoing remarks. The phenomenon (or, perhaps more accurately, family of phenomena) has in many respects been the fuel behind formal theories of truth, at least in the contemporary period. In many respects, at least in my opinion, a clear understanding of revenge is a pressing and open matter. What, exactly, is revenge? How, if at all, is it a serious problem? Is the problem *logical?* Is the problem *philosophical?* And for what end, exactly, is the alleged problem a problem? How- is it related to ECP, NTP, or the likes of ~semantic closure'?

Answers to some of the given questions, I hope, are clear enough in foregoing remarks. But answers clear answers to many of the questions remain to be found. Until then, full evaluation of current theories of truth remain out of reach.

11 FURTHER READING

This section is intended as a brief but in *no* way exhaustive suggestion towards further reading on the main topics. The bibliographies in the given works contain a broader guide. In the case of edited collections, I try to list fairly recent works, which are likely to be widely available. (Again, the bibliographies in the given works will point to many other works of direct relevance.) Unless otherwise stated, *collections* (as used here) are volumes that contain papers by various authors, rather than a single author's collected papers.

11.1 Truth and Paradox

> [Beall, 2003]. While some of the papers in this collection focus chiefly on soritical paradoxes and issues of vagueness, some of the others focus squarely on semantical or logical paradox, with a few papers tying the two kinds of paradox together.
>
> [Beall and Armour-Garb, 2005] is a collection that focuses squarely on dtruth and dtruth-theoretic paradox, and in particular on the prospects of achieving 'deflationistically acceptable' solutions to such paradox.

---

lrgFor a recent example of attempting to meet this burden, see [Field, 2005b].

[Chapuis and Gupta, 2000] is a collection that focuses on revision-theoretic approaches to paradox, and also contains discussion of extending revision theory in other philosophical and logical directions.

[Martin, 1970] is a collection that represents approaches to truth-theoretic paradox in the 1960s and early 70s.

[Martin, 1984] is a collection that represents approaches that, from the late 1970s to early 80s, have emerged as philosophically popular and influential in the area.

[Visser, 2004] is a much broader, detailed discussion of many of the approaches discussed in the current essay, with a focus on both philosophical and, especially, the mathematical aspects of the given theories. (The given bibliography is also quite valuable.)

11.2 Dtruth and 'nature' issues

[Armour-Garb and Beall, 2005] is a collection of papers representing the various deflationary conceptions of truth. (Note that 'deflationism' is really just a catch-all phrase for a family of closely related theories or, more broadly, views.)

[Blackburn and Simmons, 1999] is a collection of papers that give a flavor of the 'nature' debates. (The introductory essay pushes the relevance of truth-theoretic paradox to 'nature' questions.)

[David, 1994] is a monograph that provides an extensive examination of both correspondence theories and deflationary theories of truth.

[Greenough and Lynch, 2005] is a collection that, in many ways, ties approaches to the nature, if any, of truth to broader philosophical and methodological issues.

[Halbach and Horsten, 2002b] is a collection that discusses a wide variety of truth-related topics; it contains both 'nature' topics and discussion of truth-theoretic (and, in general, semantic) paradox.

[Kfinne, 2003] is a monograph that examines a variety of popular or historically significant positions on 'nature' issues. (Though it is not a history, the historical information in this book is valuable.)

[Lynch, 2001] contains essays on the 'nature' debates, including representatives of *correspondence theory, coherence theory, pragmatism, aletheic pluralism,* and various *deflationary* positions.

[Schantz, 2002] is a collection covering a wide range of 'nature' issues, as well as containing a few essays examining Tarski's so-called Convention T.

11.3 Paraconsistency

IPriest, 2002] is a wide-ranging survey of paraconsistent logics and their applications.

IPriest *et al.*, 2004] is a collection that focuses on dialetheic approaches to paradox and paraconsistency, more generally.

INorman *et al.*, 1989] is a collection that focuses on paraconsistency, quite generally touching on various logics and philosophical applications of such (paraconsistent) logics.

IPriest, 2001] is a monograph that discusses the motivations of dialetheism found in various branches of philosophy (and mathematics); its discussion of semantical paradox is directly relevant to Priest's version of dialetheism.

11.4 Paracomplete

As mentioned in the main discussion, paracomplete approaches are the dominant trend, at least with respect to dtruth, but also popular for other (less transparent) notions of truth. As a result, many if not all of the major works cited in the main discussion provide discussion and further references.

11.5 Tarski

Tarski's own work [1944; 1956] is quite readable (and worth reading). Discussion of Tarski's (or, at least, Tarski*'s) approach to truth-theoretic paradox may be found in almost all of the works cited in §11, but see especially the two Martin collections (see §11.1) and references therein.

11.6 Contextual

In addition to the positions sketched in the main discussion, a few other 'contextualist' positions are prominent. (And for related works, see the respective bibliographies.)

[Koons, 1992] provides a combination of %ituational truth' and Burgean 'indexical truth', with the main advantage being that only two levels are required for ~indexing', as opposed to infinitely many (as on the full Burgean proposal).

IGaiflnan, 1992] offers an approach often classified as a contextualist approach to truth. Semantic evaluation generates a 'network', and the interpretation of an utterance of ~true' depends on the place of the utterance in the given network.

[Skyrms, 1970; Skyrms, 1984] provides an 'intensional' approach to truth, wherein substitutivity of identities fails. (The overall approach is often seen as ~contextual' given the role of contextual-like sensitivity that emerges. One notable feature is that we do not get the right-left direction of standard T-biconditionals.)

11.7 Revision

Revision theorists share the basic revision-theoretic idea sketched above. Where revision theorists differ is in the details of how to extract *categorical* information from the various revision sequences. Such differences often show up as differences in how to define the revision operator at limit stages. For discussion of the main candidates and the issues that drive the differences, consult in addition to sources cited in the main discussion (above) the following works and their respective bibliographies.

[Chapuis, 1996] offers so-called *rally varied* revision sequences that are contrasted with standard alternatives.

[Chapuis and Gupta, 2000]. (See §11.1 for brief description.)

]Shapiro, 2005], while not advancing a particular revision theory, contains a valuable examination of the philosophical import of revision rules.

[Villanueva, 1997] contains essays directly on both philosophical and logical aspects of revision theory; it also contains essays on truth, in general.

[Yaqub, 1993] is a monograph advancing a distinctive set of revision rules.

11.8 Axiomatic Approaches

[Cantini, 1996] is a monograph that provides a detailed discussion of axiomatic approaches to truth (and related notions).

[Halbach and Horsten, 2002a] provides a user-friendly discussion of some of the main approaches (and surrounding issues) of axiomatic theories of truth.

[Sheard, 1994] is a user-friendly discussion of axiomatic (and, to some extent, semantic) approaches to truth.

ACKNOWLEDGEMENTS

This essay is drawn from a forthcoming monograph. I am grateful to Hartry Field, Michael Glanzberg, Daniel Nolan, Graham Priest, and Lionel Shapiro for discussion and comments. I am also grateful to the following for discussion of various topics in the essay: Tim Bays, Marion David, Chris Gauker, Jay Garfield,

Patrick Greenough, Anil Gupta, Carrie Jenkins, Michael Lynch, Doug Patterson, Stephen Read, Greg Restall, Kevin Sharp, Fritz Warfield, Crispin Wright, and also to audiences at Smith College, SUNY (Albany), University of Notre Dame, University of St Andrews, and members of the AHRC *Archd* Centre for Philosophy of Language, Logic, Mathematics, and Mind.

BIBLIOGRAPHY

[Aczel, 1988] P. Aczel. *Non-Well-Founded Sets.* Number 14 in CSLILecture Notes. CSLIPublications, Stanford, 1988.

[Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap. *Entailment: The Logic of Relevance and Necessity,* volume 1. Princeton University Press, Princeton, 1975.

[Anderson *et al.,* 1992] A. R. Anderson, N. D. Belnap, and J. M. Dunn. *Entailment: The Logic of Relevance and Necessity,* volume 2. Princeton University Press, Princeton, 1992.

[Armour-Garb and Beall, 2005] B. Armour-Garb and JC Beall, editors. *Deflationary T~'uth.* Open Court Press, Chicago, 2005.

[Barwise and Etchemendy, 1987] J. Barwise and J. Etchemendy. *The Liar.* Oxford University Press, Oxford, 1987.

[Beall and Armour-Garb, 2005] JC Beall and B. Armour-Garb, editors. *Deflationism and Paradox.* Oxford University Press, Oxford, 2005. In press.

[Beall and Restall, 2005] JC Beall and G. Restall. *Logical Pluralism.* Oxford University Press, Oxford, 2005. In press.

[Beall *et al.,* 2005] JC Beall, R. Brady, A. Hazen, G. Priest, and G. Restall. Relevant restricted quantification. *Journal of Philosophical Logic,* 2005. Forthcoming.

[Beall, 2002] JC Beall. Deflationism and gaps: untying 'not's in the debate. *Analysis,* 62(4), 2002.

[Beall, 2003] JC Beall, editor. *Liars and Heaps: New Essays on Paradox.* Oxford University Press, Oxford, 2003.

[Beall, 2005a] JC Beall. Negation's holiday: double-aspect dialetheism. In David DeVidi and Tin] Kenyon, editors, *A Logical Approach to Philosophy.* Kluwer Academic Publishers, Dordrecht, 2005. Forthcoming.

[Beall, 2005b] JC Beall. Transparent disquotationalism. In JC Beall and Brad Armour-Garb, editors, *Deflationism and Paradox,* pages 7 22. Oxford University Press, Oxford, 2005.

[Beall, 2006] JC Beall. True, false, and paranormal. *Analysis,* 2006. Forthcoming.

[Blackburn and Simmons, 1999] S. Blackburn and K. Simmons, editors. *Truth.* Oxford University Press, Oxford, 1999.

[Burge, 1979] T. Burge. Semantical paradox. *Journal of Philosophy,* 76:169 198, 1979. Reprinted in [Martin, 1984].

[Cantini, 1996] A. Cantini, editor. *Logical Frameworks for T~'uth and Abstraction: An Axiomatic Study.* Studies in Logic and the Foundations of Mathematics. Elsevier, Oxford, 1996.

[Chapuis and Gupta, 2000] A. Chapuis and A. Gupta, editors. *Circularity, Definition and Truth.* Indian Council of Philosophical Research, New Delhi, 2000.

[Chapuis, 1996] A. Chapuis. Alternative revision theories of truth. *Journal of Philosophical Logic,* 25:399 423, 1996.

[Chellas, 1980] B. F. Chellas. *Modal Logic: An Introduction.* Cambridge University Press, Cambridge, 1980.

[Curry, 1942] H.B. Curry. The inconsistencyof certain formal logics. *Journal of Symbolic Logic,* 7:115 17, 1942.

[David, 1994] M. David. *Correspondence and Disquotation.* Oxford University Press, Oxford, 1994.

[Dowden, 1984] B. Dowden. Accepting inconsistencies from the paradoxes. *Journal of Philosophical Logic,* 13:125 130, 1984.

[Dummett, 1978] M. Dummett. Truth. In M. Dummett, editor, *Truth and Other Enigmas.* Duckworth, London, 1978.

[Feferman, 1984] S. Feferman. Toward useful type-free theories, I. *Journal of Symbolic Logic,* 49:75 111, 1984. Reprinted in [Martin, 1984].

[Field, 2001] H. Field. *Truth and the Absence of Fact.* Oxford University Press, Oxford, 2001.

[Field, 2002] H. Field. Saving the t-schema from paradox. *Journal of Philosophical Logic,* 31:1 27, 2002.

[Field, 2003a] H. Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic,* 32:139 177, 2003.

[Field, 2003b] H. Field. The semantic paradoxes and the paradoxes of vagueness. In JC Beall, editor, *Liars and Heaps: New Essays on Paradox,* pages 262 311. Oxford University Press, Oxford, 2003.

[Field, 2004] H. Field. The consistency of the na~'ve (?) theory of properties. *The Philosophical Quarterly,* pages 78 104, 2004.

[Field, 2005a] H. Field. Is the liar both true and false? In JC Beall and Brad Armour-Garb, editors, *Deflationism and Paradox,* pages 23 40. Oxford University Press, Oxford, 2005. In press.

[Field, 2005b] H. Field. Solving the paradoxes, escaping revenge. 2005. To appear.

[Field, 2005c] H. Field. Variations on a theme by yablo. In JC Beall and Brad Armour-Garb, editors, *Deflationism and Paradox,* pages 53 74. Oxford University Press, Oxford, 2005. In press.

[Fine, 1975] K. Fine. Vagueness, truth and logic. *Synthese,* 30:265 300, 1975.

[Fitting, 1986] M. Fitting. Notes on the mathematical aspects of Kripke's theory of truth. *Notre Dame Journal of Formal Logic,* 27:75 88, 1986.

[Gaifman, 1992] H. Gaifman. Pointers to truth. *Journal of Philosophy,* 89:223 61, 1992.

[Glanzberg, 2001] M. Glanzberg. The liar in context. *Philosophical Studies,* 103:217 251, 2001.

[Glanzberg, 2004a] M. Glanzberg. A contextual hierarchical approach to truth and the liar paradox. *Journal of Philosophical Logic,* 33:27 88, 2004.

[Glanzberg, 2004b] M. Glanzberg. Quantification and realism. *Philosophy and Phenomenological Research,* 69:541 572, 2004.

[Glanzberg, 2004c] M. Glanzberg. Truth, reflection, and hierarchies. *Synthese,* 142:289 315, 2004.

[Greenough and Lynch, 2005] P. Greenough and M. P. Lynch. *Truth and Realism.* Oxford University Press, Oxford, 2005. In press.

[Grim, 1991] P. Grim. *The Incomplete Universe.* MIT Press, Cambridge, MA, 1991.

[Gupta and Belnap, 1993] A. Gupta and N. Belnap. *The Revision Theory of Truth.* MIT Press, 1993.

[Gupta, 1982] A. Gupta. Truth and paradox. *Journal of Philosophical Logic,* 11:1 60, 1982. Reprinted in [Martin, 1984].

[Hajek *et al.,* 2000] P. Hajek, J. Paris, and J. Shepherdson. The liar paradox and fuzzy logic. *Journal of Symbolic Logic,* 65:339 346, 2000.

[Halbach and Horsten, 2002a] V. Halbach and L. Horsten. Contemporary methods for investigating the concept of truth. In Volker Halbach and Leon Horsten, editors, *Principles of Truth,* Epistemische Studien. Ontos Verlag, 2002.

[Halbach and Horsten, 2002b] V. Halbach and L. Horsten, editors. *Principles of Truth.* Epistemische Studien. Ontos Verlag, 2002.

[Harman, 1986] G. Harman. *Change In View: Principles of Reasoning.* Bradford Books. MIT Press, 1986.

[Herzberger, 1982] H. Herzberger. Notes on naive semantics. *Journal of Philosophical Logic,* 11:61 102, 1982. Reprinted in [Martin, 1984].

[Hyde, 1997] D. Hyde. From heaps and gaps to heaps of gluts. *Mind,* 106:641 660, 1997.

[Jacquette, 2004] D. Jacquette. Diagonalization in logic and mathematics. In Dov M. Gabbay and Franz Giinthner, editors, *Handbook of Philosophical Logic,* pages 55 147. Kluwer Academic Publishers, Dordrecht, second edition, 2004.

[Kaplan, 1989] D. Kaplan. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes from Kaplan,* pages 481 614. Oxford University Press, New York, 1989.

[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics.* North-Holland, 1952.

[Koons, 1992] R. C. Koons. *Paradoxes of Belief and Strategic Rationality.* Cambridge University Press, Cambridge, 1992.

[Kremer, 1988] M. Kremer. Kripke and the logic of truth. *Journal of Philosophical Logic,* 17:225 278, 1988.

[Kremer, 2000] P. Kremer. 'semantics' for languages with their own truth predicates. In Andr6 Chapuis and Anil Gupta, editors, *Circularity, Definition and Truth,* pages 217 46. Indian Council of Philosophical Research, New Delhi, 2000.

[Kripke, 1965] S. A. Kripke. Semantical analysis of modal logic II: Non-normal modal propositional calculi. In Addison et al., editor, *The Theory of Models,* pages 206 220. North-Holland Publishing Co, 1965.

[Kripke, 1975] S. Kripke. Outline of a theory of truth. *Journal of Philosophy,* 72:690 716, 1975. Reprinted in]Martin, 1984].

[Kfinne, 2003] W. Kiinne. *Conceptions of Truth.* Oxford University Press, New York, 2003.

[Lewis and Langford, 1932] C. I. Lewis and C. H. Langford. *Symbolic Logic.* The Century Co, New York and London, 1932.

[Lynch, 1998] M. P. Lynch. *Truth in Context.* MIT Press, Cambridge, MA, 1998.

[Lynch, 2001] M. P. Lynch, editor. *The Nature of Truth.* MIT Press, Cambridge, MA, 2001.

[Mares, 2004] E. D. Mares, editor. *Relevant Logic: A Philosophical Interpretation.* Cambridge University Press, Cambridge, 2004.

[Martin and Woodruff, 1975] R. L. Martin and P. W. Woodruff. On representing 'true-in-L' in *L. Philosophia,* 5:217 221, 1975. Reprinted in [Martin, 1984].

[Martin, 1970] R. L. Martin, editor. *The Liar Paradox.* Yale University Press, New Haven, 1970.

[Martin, 1984] R. L. Martin, editor. *Recent Essays on Truth and the Liar Paradox.* Oxford University Press, New York, 1984.

[Maudlin, 2004] T. Maudlin. *Truth and Paradox: Solving the Riddles.* Oxford University Press, New York, 2004.

[McGee, 1991] V. McGee. *Truth, Vagueness, and Paradox.* Hackett, Indianapolis, 1991.

[Meyer *et al.,* 1979] R. K. Meyer, R. Routley, and J. M. Dunn. Curry's paradox. *Analysis,* 39:124 128, 1979.

[Norman *et al.,* 1989] J. Norman, G. Priest, and R. Sylvan, editors. *Paraconsistent Logic: Essays on the Inconsistent.* Philosophia Verlag, 1989.

[Parsons, 1974a] C. Parsons. The liar paradox. *Journal of Philosophical Logic,* 3:381 412, 1974. Reprinted in]Martin, 1984] and]Parsons, 1983].

[Parsons, 1974b] C. Parsons. Sets and classes. *No~ts,* 8:1 12, 1974. Reprinted in [Parsons, 1983].

[Parsons, 1983] C. Parsons. *Mathematics in Philosophy.* Cornell University Press, Ithica, 1983.

[Parsons, 1990] T. Parsons. True contradictions. *Canadian Journal of Philosophy,* 20:335 353, 1990.

[Patterson, 2005] D. E. Patterson. Tarski, the liar, and inconsistent languages. *The Monist,* 89(1), 2005. In press.

[Priest *et al.,* 2004] G. Priest, JC Beall, and B. Armour-Garb, editors. *The Law of Non-Contradiction.* Oxford University Press, Oxford, 2004.

[Priest, 1979] G. Priest. The logic of paradox. *Journal of Philosophical Logic,* 8:219 241, 1979.

[Priest, 1987] G. Priest. *In Contradiction: A Study of the Transconsistent.* Martinus Nijhoff, The Hague, 1987.

[Priest, 1992] G. Priest. What is a non-normal world? *Logique et Analyse,* 35:291 302, 1992.

[Priest, 1998] G. Priest. What is so bad about contradictions? *Journal of Philosophy,* XCV(8):410 426, 1998. Reprinted in]Priest *et al.,* 2004].

[Priest, 2000] G. Priest. Motivations for paraconsistency: The slippery slope from classical logic to dialetheism. In D. Batens, C. Mortensen, G. Priest, and J-P Van Bendegem, editors, *Frontiers of Paraconsistency.* Research Studies Press, KingOs College Publications, Baldock, 2000.

[Priest, 2001] G. Priest. *Beyond the Limits of Thought.* Oxford University Press, Oxford, second edition, 2001.

[Priest, 2002] G. Priest. Paraconsistent logic. In Dov M. Gabbay and Franz Gfinthner, editors, *Handbook of Philosophical Logic (2nd Edition),* volume 6, pages 287 393. Kluwer Academic Publishers, Dordrecht, 2002.

[Priest, 2005] G. Priest. Spiking the Field artillery. In JC Beall and Brad Armour-Garb, editors, *Deftationisrn and Paradox,* pages 41 52. Oxford University Press, Oxford, 2005. In press.

[Priest, 2006a] G. Priest. *Doubt Truth To Be A Liar'.* Oxford University Press, Oxford, 2006. In press.

[Priest, 2006b] G. Priest. *In Contradiction.* Oxford University Press, Oxford, second edition, 2006. Forthcoming.

[Quine, 1970] W. V. O. Quine. *Philosophy of Logic.* Prentice-Hall, 1970.

[Ramsey, 1925] F. P. Ramsey. The foundations of mathematics. *Proceedings of the London Mathematical Society,* 25:338 384, 1925.

[Reinhardt, 1986] W. N. Reinhardt. Some remarks on extending and interpreting theories with a partial predicate for truth. *Journal of Philosophical Logic,* 15:219 251, 1986.

[Rescher and Brandom, 1979] N. Rescher and R. Brandom. *The Logic of Inconsistency.* Rowman and Littlefield, Totowa, NJ, 1979.

[Restall, 1992] G. Restall. Arithmetic and truth in Lukasiewicz's infinitely valued logic. *Logique et AnalysG* 139 140:303 312, 1992.

[Restall, 2000] G. Restall. *An Introduction to Substructural Logics.* New York: Routledge, 2000.

[Routley and Loparic, 1978] R. Routley and A. Loparic. Semantical analyses of Arruda-daCosta P-systems and adjacent non-replacement systems. *Studia Logica,* 37:301 320, 1978.

[Routley *et al.,* 1982] R. Routley, V. Plumwood, R. K. Meyer, and R. T. Brady, editors. *Relevant Logics and their Rivals.* Ridgeview, 1982.

[Schantz, 2002] R. Schantz, editor. *What is Truth?* Walter de Gruyter, New York, 2002.

[Shapiro, 2004] S. Shapiro. Simple truth, contradiction, and consistency. In Graham Priest, JC Beall, and Brad Armour-Garb, editors, *The Law of Non-Contradiction.* Oxford University Press, 2004.

]Shapiro, 2005] L. Shapiro. The rationale behind revision-rule semantics. *Philosophical Studies,* 2005. Forthcoming.

[Shaw-Kwei, 1954] M. Shaw-Kwei. Logical paradoxes for many-valued systems. *Journal of Symbolic Logic,* 19:37 40, 1954.

[Sheard, 1994] M. Shear& A guide to truth predicates in the modern era. *Journal of Symbolic Logic,* 59(3):1032 54, 1994.

[Simmons, 1993] K. Simmons. *Universality and The Liar.* Cambridge University Press, 1993.

[Skyrms, 1970] B. Skyrms. Return of the liar: Three-valued logic and the concept of truth. *American Philosophical Quarterly,* 7:153 161, 1970.

[Skyrms, 1984] B. Skyrms. Intensional aspects of semantical self-reference. In R. M. Martin, editor, *Recent Essays on Truth and the Liar,* pages 119 131. Oxford University Press, Oxford, 1984.

[Smullyan, 1992] R. M. Smullyan. *GSdel's Incompleteness Theorems,* volume 19 of *Oxford Logic Guides.* Oxford University Press, New York, 1992.

[Smullyan, 1993] R.M. Smullyan. *Recursion Theory for Metamathematics,* volume 20 of *Oxford Logic Guides.* Oxford University Press, New York, 1993.

[Soames, 1999] S. Soames. *Understanding Truth.* Oxford University Press, New York, 1999.

[Tarski, 1936] A. Tarski. Der wahreitsbegriff in den formalisierten sprachen. *Studia Philosophica,* 1:261 405, 1936. Reprinted in English in [Tarski, 1956].

[Tarski, 1944] A. Tarski. The semantic concept of truth and the foundations of semantics. *Philosophy and Phenomenological Research,* 4:341 376, 1944.

[Tarski, 1956] A. Tarski. *Logic, Semantics, Metamathematics: papers from 1923 to 1938.* Clarendon Press, Oxford, 1956. Translated by J. H. Woodger.

[van Fraassen, 1968] B. C. van Fraassen. Presupposition, implication and self-reference. *Journal of Philosophy,* 65:136 52, 1968.

[van Fraassen, 1970] B. C. van Fraassen. Truth and paradoxical consequences. In R. Martin, editor, *The Paradox of the Liar,* pages 13 23. Yale University Press, 1970.

[Varzi, 1999] A. Varzi. *An Essay in Universal Semantics,* volume 1 of *Topoi Library.* Kluwer Academic Publishers, Boston, 1999.

[Villanueva, 1997] E. Villanueva, editor. *Truth.* Number 8 in Philosophical Issues. Ridgeview Publishing Company, Atascadero, California, 1997.

[Visser, 1984] A. Visser. Four valued semantics and the liar. *Journal of Philosophical Logic,* 13:181 212, 1984.

[Visser, 2004] A. Visser. Semantics and the liar paradox. In Dov M. Gabbay and Franz Giinthner, editors, *Handbook of Philosophical Logic,* pages 149 240. Kluwer Academic Publishers, Dordrecht, second edition, 2004.

[White, 1979] R. B. White. The consistency of the axiom of comprehension in the infinite-valued predicate logic of tukasiewicz. *Journal of Philosophical Logic,* 8:509 534, 1979.

[Woodruff, 1984] P. Woodruff. Paradox, truth, and logic. *Journal of Philosophical Logic,* 13:213 252, 1984.

[Wright, 1987] C. Wright. *Realism, meaning and truth.* Oxford University Press, 1987.

[Wright, 1992] C. Wright. *Truth and Objectivity.* Harvard University Press, 1992.

[Yab]o, 1993a] S. Yablo. Definitions, consistent and inconsistent. *Philosophical Studies,* 72:147 175, 1993. Special Issue on Definition.

[Yab]o, 1993b] S. Yablo. Hop, skip, and jump. In James Tomberlin, editor, *Philosophical Perspectives: Language and Logic,* volume 7, pages 371 396. Ridgeview Publishing Co., Atascadero, CA, 1993.

[Yab]o, 2003] S. Yablo. New grounds for na'fve truth theory. In JC Beall, editor, *Liars and Heaps: New Essays on Paradox,* pages 313 330. Oxford University Press, Oxford, 2003.

[Yaqfib, 1993] A. Yaqfib. *The Liar @eaks the Truth.* Oxford University Press, 1993.

HILBERT'S PROGRAM THEN AND NOW

Richard Zach

1 INTRODUCTION

Hilbert's program is, in the first instance, a proposal and a research program in the philosophy and foundations of mathematics. It was formulated in the early 1920s by German mathematician David Hilbert (1862–1943), and was pursued by him and his collaborators at the University of Göttingen and elsewhere in the 1920s and 1930s. Briefly, Hilbert's proposal called for a new foundation of mathematics based on two pillars: the axiomatic method and finitary proof theory. Hilbert thought that by formalizing mathematics in axiomatic systems, and subsequently proving by finitary methods that these systems are consistent (i.e., do not prove contradictions), he could provide a philosophically satisfactory grounding of classical, infinitary mathematics (analysis and set theory). Had it been successful, Hilbert's program would perhaps not only have eclipsed in subsequent influence other foundational projects of the time, such as the logicist projects pursued by Frege and Russell and Brouwer's intuitionism, but it would also have achieved Hilbert's stated goal, viz., to "dispose of the foundational questions in mathematics as such once and for all" [Hilbert, 1929, 228]. Unfortunately, Gödel's theorems show that the program as originally envisaged by Hilbert cannot be carried out.

Although Hilbert's own project for the foundation of mathematics was ultimately unsuccessful, the project itself and technical advances made in its pursuit have had an enormous impact on logic and the foundations of mathematics more generally. In order to carry out the first part of the program, the axiomatization of mathematics in a formal system, Hilbert and his collaborators had pushed forward the development of logical formalisms in which such an axiomatization could be carried out. This led, in particular, to the first axiomatizations of propositional and first-order logic as independent systems (i.e., other than as fragments of more comprehensive higher-order systems, such as Frege's *Begriffsschrift* and the Whitehead-Russell system of *Principia mathematica*) and their metalogical investigation, such as the proof (by Paul Bernays in 1918) of the completeness of the propositional calculus and the work in Hilbert's school from 1921 onward on the decision problem. The investigation of the metalogical properties of logical systems led directly to some of the most important metalogical results in logic, viz., Gödel's completeness theorem and the negative solution by Church and Turing of the decision problem. The development of proof theory itself is an outgrowth of Hilbert's program. Gentzen's development of natural deduction and the sequent

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

calculus was carried out in the tradition of Hilbert's program and with the aim of constructing a logical system which facilitates consistency proofs. Gödel obtained his incompleteness theorems while trying to prove the consistency of analysis. And the tradition of reductive proof theory of the Gentzen-Schütte school and others is itself a direct continuation of Hilbert's program.

The present chapter is divided into three parts: The first part provides a sketch of the historical development of logic, proof theory, and philosophy of mathematics in the work of Hilbert and his followers through the 1930s. The second part deals with the philosophical interpretation and assessment of Hilbert's program. The third part presents recent work in proof theory which bears on the aims of Hilbert's program.

2 HILBERT'S PROGRAM THEN

2.1 Hilbert's early work on foundations

Hilbert's work on the foundations of mathematics can be traced to his work on geometry of the 1890s which resulted in his influential textbook *Foundations of Geometry* [1899]. One philosophical advance of this work was the development of Hilbert's conception of the axiomatic method. Hilbert believed that the proper way to develop any scientific subject rigorously required an axiomatic approach. In providing an axiomatic treatment, the theory would be developed independently of any need for intuition, and it would facilitate an analysis of the logical relationships between the basic concepts and the axioms. Of basic importance for an axiomatic treatment are, so Hilbert, investigation of the independence and, above all, of the consistency of the axioms. In his 1902 lectures on the foundations of geometry, he puts it thus:

> Every science takes its starting point from a sufficiently coherent body of facts as given. It takes form, however, only by *organizing* this body of facts. This organization takes place through the *axiomatic method*, i.e., one constructs a *logical structure of concepts* so that the relationships between the concepts correspond to relationships between the facts to be organized.
>
> There is arbitrariness in the construction of such a structure of concepts; we, however, *demand* of it:
>
> 1) completeness, 2) independence, 3) consistency. [Hilbert, 2004, 540]

From the time of his work on geometry forward, the last consideration, consistency, was of special importance in Hilbert's conception of the axiomatic method in general and the foundations of mathematics in particular. Hilbert was heavily influenced by the foundational views of late-19th century mathematicians, in particular, Cantor, Dedekind, and Kronecker. He shared with Dedekind and Cantor the view that mathematical activity should be free of constraints, which led to his view, highlighted in his correspondence with Frege, that consistency of an axiomatic theory guarantees the existence of the structure described, and is in this sense sufficient to justify the use of the theory. And he shared with Kronecker a recognition that elementary arithmetic has a privileged role in mathematics, although he was of course opposed to the converse espoused by Kronecker, viz., that the natural numbers, and constructions based on elementary arithmetic, exhaust legitimate mathematics. These two influences in Hilbert's thought are at the root of his investigations of consistency.

Proofs of consistency for the axioms of geometry can be given by providing an interpretation of the system in the real plane, and thus the consistency of geometry is reduced to the consistency of analysis. Analysis, of course, itself requires justification. In [1900b], Hilbert approached the problem from the axiomatic standpoint by proposing an axiomatization of the real numbers. In order to show the consistency of this system, Hilbert expressly rejected the construction of a model, e.g., a construction based on Dedekind cuts of rationals, as an option. He considered the construction of the reals from the rationals and ultimately the natural numbers using the "genetic method" as insufficient: "Despite the high pedagogical and heuristic value of the genetic method, for the final presentation and the complete logical grounding of our knowledge the axiomatic method deserves the first rank" [Hilbert, 1900b, 1093]. Hilbert thus was after a *direct* consistency proof of analysis, i.e., one not based on reduction to another theory. He proposed the problem of finding such a proof as the second of his 23 mathematical problems in his address to the International Congress of Mathematicians in 1900 [1900a].

The discovery of Russell's paradox in 1902 made it clear that an axiomatic foundation of arithmetic and set theory requires a more precise development of the underlying logical systems. Hilbert knew of the paradoxes of set theory from Cantor and Zermelo, but it was apparently not until Russell's [1902] publication of the contradiction in Frege's system that Hilbert and Zermelo realized their importance. Hilbert's exchange with Frege on the axiomatic approach to geometry led him to realize that his conceptions of "axiom," "definition," "proof" were in need of clarification. In response, Hilbert formulated an early version of his proof-theoretical program in his 1904 Heidelberg talk [1905]. After criticizing the foundational views of Kronecker as dogmatic, and those of Frege and Dedekind as suffering from "unavoidable contradictions," he writes:

> Arithmetic is often considered to be a part of logic, and the traditional fundamental logical notions are usually presupposed when it is a question of establishing a foundation for arithmetic. If we observe attentively, however, we realize that in the traditional exposition of the laws of logic certain fundamental arithmetic notions are already used, for example, the notion of set and, to some extent, also that of number. Thus we find ourselves turning in a circle, and that is why a partly simultaneous development of the laws of logic and of arithmetic is required if paradoxes are to be avoided. [Hilbert, 1905, 131]

Hilbert's sketch of this "simultaneous development" of logic and arithmetic in the case of a very basic axiom system for the natural numbers is very close to the approach Hilbert's proof theoretic program would take 20 years later: Hilbert gives a direct argument that no contradiction can arise from the five axioms of his system.

This was a promising start, but several factors delayed the further development of Hilbert's proof theoretic program. One was Poincaré's [1906] criticism of what he saw as a viciously circular use of induction in Hilbert's sketched consistency proof (see [Steiner, 1975], Appendix). Moreover, Hilbert realized that axiomatic investigations required a well worked-out logical formalism in which axiomatic systems could be developed. At the time he used a logical formalism based on Schröder's algebraic logic, which was not particularly suited as a formalism for the axiomatization of mathematics. Following the 1905 lectures on foundations, Hilbert turned his immediate attention to work in other areas of mathematics and theoretical physics. He did, however, actively support others who worked on foundational questions in Göttingen, in particular Ernst Zermelo and Leonard Nelson.[1]

The publication of Whitehead and Russell's *Principia Mathematica* [1910; 1912; 1913] provided the required logical basis for a renewed attack on foundational issues. Beginning in 1914, Hilbert's student Heinrich Behmann and others studied the system of *Principia*.[2] Hilbert himself returned to work on the foundations of mathematics in 1917. In September 1917, he delivered an address to the Swiss Mathematical Society entitled "Axiomatic Thought" [1918a]. It is his first published contribution to mathematical foundations since 1905. In it, he again emphasized the requirement of consistency proofs for axiomatic systems: "The chief requirement of the theory of axioms must go farther [than merely avoiding known paradoxes], namely, to show that within every field of knowledge contradictions based on the underlying axiom-system are *absolutely impossible*." He posed the proof of the consistency of the integers (and of set theory) again as the main problems. In both these cases, there seems to be nothing more fundamental available to which the consistency could be reduced other than logic itself. Hilbert at the time considered the problem as essentially solved by Whitehead and Russell's work in *Principia*. Nevertheless, other fundamental problems of axiomatics remained unsolved, including the problem of the "decidability of every mathematical question," which also traces back to Hilbert's 1900 address.

---

[1] See [Sieg, 1999; Sieg, 2002], [Stein, 1988], [Hallett, 1990; Hallett, 1994], [Mancosu, 1998b], and [Avigad and Reck, 2001] for further discussion of the conceptual framework and historical background of Hilbert's thought, and [Resnik, 1974a] on the Frege-Hilbert correspondence. On Hilbert's foundational interests before 1917, and his engagement for Husserl, Zermelo, and Nelson in Göttingen, see [Peckhaus, 1990]. On general discussions of formalism and the place of Hilbert's thought in the mathematical context of the late 19th century, see [Webb, 1997] and [Detlefsen, 2005].

[2] See [Mancosu, 1999] and [2003] on Behmann's role in Hilbert's school and the influence of Russell.

These unresolved problems of axiomatics led Hilbert to devote significant effort to work on logic in the following years. In 1917, Paul Bernays joined him as his assistant in Göttingen. In a series of courses from 1917–1921, Hilbert, with the assistance of Bernays and Behmann, made significant new contributions to formal logic. The course from 1917 [Hilbert, 1918b], in particular, contains a sophisticated development of first-order logic, and forms the basis of Hilbert and Ackermann's textbook *Principles of Theoretical Logic* [1928]. In 1918, Bernays submitted a treatise on the propositional calculus of *Principia mathematica* as a *Habilitationsschrift*; it contains the first completeness proof of the propositional calculus for truth-functional semantics.[3]

The 1917–18 lectures were only the beginning of a strand of work on logic and metalogic in Hilbert's school, including work on the decision problem. The decision problem for first-order logic was tightly bound up with the aim of finding a completeness proof for the first-order predicate calculus (the "restricted calculus of functions" in Hilbert's terminology). This aim was stated in the 1917–18 lectures, but since completeness does not hold for first-order logic in any purely syntactic sense (an early result due to Ackermann), a development of the semantics of first-order logic was needed first. The decision problem, one of Hilbert's main aims for metamathematics in the 1920s, was already an issue in the lectures from 1905, and has its roots in Hilbert's belief, first explicitly stated in the Paris address, that "in mathematics, there is no ignorabimus," i.e., that every mathematical question can be solved either affirmatively or negatively. The questions of completeness and decidability thus became closely linked in the 1920s, with Behmann, Bernays, Schönfinkel, and later Ackermann working on special cases of the decision problem for first-order logic throughout the 1920s.

2.2 The consistency program, finitism, and proof theory

In about 1920, Hilbert came to reject Russell's logicist solution to the consistency problem for arithmetic, mainly for the reason that the axiom of reducibility cannot be accepted as a purely logical axiom. In lectures from the Summer term 1920, he concluded that "the aim of reducing set theory, and with it the usual methods of analysis, to logic, has not been achieved today and maybe cannot be achieved at all" [Hilbert, 1920]. At the same time, Brouwer's intuitionist mathematics gained currency. In particular, Hilbert's former student Hermann Weyl converted to intuitionism. Weyl's 1920 address to the Hamburg Mathematical Seminar, "The new foundational crisis in mathematics" [1921] was answered by Hilbert in three talks in Hamburg in the Summer of 1921 [1922b]. Here, Hilbert presented his own mature proposal for a solution to the problem of the foundation of mathematics. This proposal incorporated Hilbert's ideas from 1904 regarding direct consistency proofs, his conception of axiomatic systems, and also the technical developments in the axiomatization of mathematics in the work of Russell as well as the further developments carried out by him and his collaborators. What was new was the way in which Hilbert wanted to imbue his consistency project with the philosophical significance necessary to answer Brouwer and Weyl's criticisms: the finitary point of view.

---

[3] See [Moore, 1997], [Sieg, 1999] and [Zach, 1999] for more detail on the development of logic in Hilbert's school around 1918.

According to Hilbert, there is a privileged part of mathematics, contentual elementary number theory, which relies only on a "purely intuitive basis of concrete signs." Whereas the operating with abstract concepts was considered "inadequate and uncertain," there is a realm of

> extra-logical discrete objects, which exist intuitively as immediate experience before all thought. If logical inference is to be certain, then these objects must be capable of being completely surveyed in all their parts, and their presentation, their difference, their succession (like the objects themselves) must exist for us immediately, intuitively, as something which cannot be reduced to something else.[4]

The objects in questions are *signs*, both numerals and the signs that make up formulas and formal proofs. The domain of contentual number theory consists in the finitary numerals, i.e., sequences of strokes. These have no meaning, i.e., they do not stand for abstract objects, but they can be operated on (e.g., concatenated) and compared. Knowledge of their properties and relations is intuitive and unmediated by logical inference. Contentual number theory developed this way is secure, according to Hilbert: no contradictions can arise simply because there is no logical structure in the propositions of contentual number theory.

The intuitive-contentual operations with signs form the basis of Hilbert's metamathematics. Just as contentual number theory operates with sequences of strokes, so metamathematics operates with sequences of symbols (formulas, proofs). Formulas and proofs can be syntactically manipulated, and the properties and relationships of formulas and proofs are similarly based in a logic-free intuitive capacity which guarantees certainty of knowledge about formulas and proofs arrived at by such syntactic operations. Mathematics itself, however, operates with abstract concepts, e.g., quantifiers, sets, functions, and uses logical inference based on principles such as mathematical induction or the principle of the excluded middle. These "concept-formations" and modes of reasoning had been criticized by Brouwer and others on grounds that they presuppose infinite totalities as given, or that they involve impredicative definitions (which were considered by the critics as viciously circular). Hilbert's aim was to justify their use. To this end, he pointed out that they can be formalized in axiomatic systems (such as that of *Principia* or those developed by Hilbert himself), and mathematical propositions and proofs thus turn into formulas and derivations from axioms according to strictly circumscribed rules of derivation. Mathematics, so Hilbert, "becomes an inventory of provable formulas." In this way the proofs of mathematics are subject to metamathematical, contentual investigation. The goal of Hilbert's program is then to give a contentual, metamathematical proof that there can be no derivation of a contradiction, i.e., no formal derivation of a formula A and of its negation $\neg A$.

---

[4] [Hilbert, 1922b, 202]. The passage is repeated almost verbatim in [Hilbert, 1926, 376], [Hilbert, 1928, 464], and [Hilbert, 1931b, 267]

This sketch of the aims of the program was fleshed out by Hilbert and his collaborators in the following 10 years. On the conceptual side, the finite standpoint and the strategy for a consistency proof were elaborated by Hilbert [1923; 1926; 1928] and Bernays [1922; 1928b; 1930], of which Hilbert's article "On the infinite" [1926] provides the most detailed discussion of the finitary standpoint. In addition to Hilbert and Bernays, a number of other people were involved in technical work on the program. The ε-operator was first introduced in the Hamburg lecture of 1921 [Hilbert, 1922b], and developed in a number of lectures given in Göttingen [Hilbert, 1922a; Hilbert and Bernays, 1923], as well as in [Hilbert, 1923]. Hilbert and Bernays developed the ε-calculus as their definitive formalism for axiom systems for arithmetic and analysis, and the so-called ε-substitution method as the preferred approach to giving consistency proofs.

Briefly, the ε-calculus is a formalism that includes ε as a term-forming operator. If $A(x)$ is a formula, then $\varepsilon_x A(x)$ is a term, which intuitively stands for a witness for $A(x)$. In a logical formalism containing the ε-operator, the quantifiers can be defined by: $\exists x\, A(x) \equiv A(\varepsilon_x A(x))$ and $\forall x\, A(x) \equiv A(\varepsilon_x \neg A(x))$. The only additional axiom necessary is the so-called "transfinite axiom," $A(t) \rightarrow A(\varepsilon_x A(x))$. Based on this idea, Hilbert and his collaborators developed axiomatizations of number theory and analysis. Consistency proofs for these systems were then given using the ε-substitution method. The idea of this method is, roughly, that the ε-terms $\varepsilon_x A(x)$ occurring in a formal proof are replaced by actual numerals, resulting in a quantifier-free proof. The simplest case, outlined in Hilbert's papers, is as follows. Suppose we had a (suitably normalized) derivation of $0 = 1$ that contains only one ε-term $\varepsilon_x A(x)$. Replace all occurrences of $\varepsilon_x A(x)$ by 0. The instances of the transfinite axiom then are all of the form $A(t) \rightarrow A(0)$. Since no other ε-terms occur in the proof, $A(t)$ and $A(0)$ are basic numerical formulas without quantifiers and, we may assume, also without free variables. So they can be evaluated by finitary calculation. If all such instances turn out to be true numerical formulas, we are done. If not, this must be because $A(t)$ is true for some t, and $A(0)$ is false. Then replace $\varepsilon_x A(x)$ instead by n, where n is the numerical value of the term t. The resulting proof is then seen to be a derivation of $0 = 1$ from true, purely numerical formulas using only modus ponens, and this is impossible. Indeed, the procedure works with only slight modifications even in the presence of the induction axiom, which in the ε-calculus takes the form of a least number principle: $A(t) \rightarrow \varepsilon_x A(x) \leq t$, which intuitively requires $\varepsilon_x A(x)$ to be the *least* witness for $A(x)$.

The ε-substitution method is simple enough for the basic cases considered by Hilbert, but becomes extremely complex when ε-operators are nested. In his 1924 dissertation, [Ackermann, 1925] presented an (erroneous) consistency proof based on the ε-substitution method for a version of analysis. John von Neumann, then visiting Göttingen, gave a corrected consistency proof for a system of the ε-formalism (which, however, did not include the induction axiom) in 1925 [1927].

Building on von Neumann's work, Ackermann devised a new ε-substitution procedure which he communicated to Bernays (see [Bernays, 1928b]). Ackermann and Bernays considered the proof to be correct for the entire first-order fragment of arithmetic and were confident that it could be extended to a consistency proof of analysis. In his address "Problems of the grounding of mathematics" to the International Congress of Mathematicians in Bologna in 1928 [1929], Hilbert optimistically claimed that the work of Ackermann and von Neumann had established the consistency of number theory and that the proof for analysis had already been carried out by Ackermann "to the extent that the only remaining task consists in the proof of an elementary finiteness theorem that is purely arithmetical."[5]

2.3 The impact of Gödel's incompleteness theorems

Gödel's incompleteness theorems [Gödel, 1931] showed that Hilbert's optimism was unfounded. In September 1930, Kurt Gödel announced his first incompleteness theorem at a conference in Königsberg. Von Neumann, who was in the audience, immediately recognized the significance of Gödel's result for Hilbert's program. Shortly after the conference he wrote to Gödel, telling him that he had found a corollary to Gödel's result. Gödel had found the same result already independently: the second incompleteness theorem, asserting that the system of *Principia* does not prove the formalization of the claim that the system of *Principia* is consistent (provided it is). All the methods of finitary reasoning used in the consistency proofs up till then were believed to be formalizable in *Principia*, however. Hence, if the consistency of *Principia* were provable by the methods used in Ackermann's proofs, it should be possible to formalize this proof in *Principia*; but this is what the second incompleteness theorem states is impossible. Bernays also immediately realized the importance of Gödel's results after he studied Gödel's paper in January 1931. He wrote to Gödel that (under the assumption that finitary reasoning can be formalized in *Principia*) the incompleteness theorems seem to show that a finitary consistency proof of *Principia* is impossible. Shortly thereafter, von Neumann showed that Ackermann's consistency proof is flawed and provided a counterexample to the proposed ε-substitution procedure.[6]

Although the impact of Gödel's incompleteness theorems for Hilbert's program was recognized soon after their publication, Hilbert's program was by no means abandoned. Hilbert himself was no longer actively involved in foundational research, but Bernays continued to work on the foundations of mathematics. The two-volume *Grundlagen der Mathematik* [Hilbert and Bernays, 1934; Hilbert and Bernays, 1939] was prepared by Bernays alone, and included new results by Ackermann and Bernays on the ε-calculus. It also included Herbrand's [1930] work on proof theory, and a sketch of Gentzen's [1936] consistency proof of first-order arithmetic. Bernays's and Gentzen's work, in particular, focused on possible extensions of the finitary standpoint which could yield consistency proofs for substantial parts of mathematics in spite of Gödel's theorems.

---

[5] See [Avigad and Zach, 2002] for a general introduction to the ε-calculus and [Zach, 2003b] and [2004] on the history of the ε-calculus and the substitution method. [Sieg, 1999] presents a detailed and perceptive analysis of the development of the program and its influence as a whole.

[6] The correspondence between Gödel and Bernays is published in [Gödel, 2003a, 41–313] and that with von Neumann in [Gödel, 2003b, 327–377]. See also the informative introductions by Feferman and Sieg, respectively, to these sections of the correspondence, as well as [Mancosu, 2004] and the last section of [Zach, 2003b].

Gentzen's first consistency proof for number theory, the so-called galley proof [1935], was the result of a combination of Gentzen's [1934] earlier work on the logical formalism of the sequent calculus, which provided a proof-theoretically more convenient axiomatization of arithmetic, and a new strategy of proving consistency. This strategy involved defining certain "reduction steps" on proofs: local transformations of parts of a derivation in the new formalism. The consistency proof then proceeds by showing that by iterating these reductions on a proof one eventually arrives at a proof of a special form (a proof free of the cut rule and the induction rule), and no proof of such a form can be the proof of a contradiction. The first version of the proof relied on the notion of a reduction rule, which itself cannot be formalized in arithmetic.[7] This notion met with some objections, and in the revised, published version [1936], Gentzen replaced the appeal to reduction rules by a proof that the iteration of reduction steps itself terminates. He did this by assigning a measure for the complexity of a given derivation, and showed that the result of the application of a reduction step to a proof reduces the complexity measure of that proof. The complexity measure Gentzen used was certain finite strings of numbers which may be interpreted as naming countable ordinals less than ε_0.[8] The consistency result then follows if one accepts that there are no infinite descending sequences of such ordinal notations, or, more precisely, by using transfinite induction up to ε_0. This principle, by Gödel's incompleteness theorem, cannot itself be formalized in first-order arithmetic [Gentzen, 1943]. Gentzen's proof allowed Ackermann [1940] to give a correct consistency proof based on the ε-substitution method for first-order arithmetic, also based on transfinite induction up to ε_0.

Gentzen's work marks the beginning of post-Gödelian proof theory. In the proof-theoretic tradition of Gentzen, axiomatic theories are analyzed according to which transfinite induction principles are required to prove the consistency of the theory. However, Gentzen's contribution and influence goes beyond this: He emphasized that proof-theoretic methods do not just allow us to prove the consistency of a theory, but that they also enable us to extract information from proofs beyond the fact that the formula proved follows from the axioms.

3 PHILOSOPHICAL INTERPRETATION OF HILBERT'S PROGRAM

The philosophical importance and influence of Hilbert's work on foundations is twofold. First, the epistemological standpoint of Hilbert's finitism is of interest in the philosophy of mathematics quite independently of the success of the proof-theoretic program which it underlies. Like other important proposals in the philosophy of mathematics such as intuitionism, predicativism, and logicism, Hilbert's finitism is, *inter alia*, a philosophical view about the nature of mathematical knowledge and delineates a particular set of (finitarily) meaningful propositions, as well as finitarily admissible constructions and methods of proof. Debate about the status of finitary evidence and proof are still very much alive today. Second, Hilbert's program can and has been seen as a version of reductive instrumentalism in mathematics. That is to say, one can read Hilbert as proposing that only a certain part of mathematics (propositions, proofs) is meaningful, viz., the finitary part. The rest, which includes classical infinitary mathematics (full first-order arithmetic, analysis, and set theory, in particular) are mere instruments. Hilbert's program has thus been an important inspiration for related instrumentalist proposals in the philosophy of mathematics (e.g., [Field, 1980; Detlefsen, 1986]).

---

[7] On the galley proof, see [Bernays, 1970], [Kreisel, 1971, Appendix II], and [Negri, 1980].

[8] If $\omega_0 = \omega$, and $\omega_{n+1} = \omega^{\omega_n}$, then the ordinal ε_0 is the limit of ω_n for $n = 1, 2, \ldots$. In other words, ε_0 is the least fixed point of $\alpha = \omega^\alpha$.

This section will expand on these two themes. In the case of both the debate on the philosophy of finitism, and on the view of Hilbert's program as an instrumentalist philosophy of mathematics, questions of historical interpretation interact with conceptual analysis. The distinction between these aspects should be kept in mind.

3.1 The finitary point of view

The cornerstone of Hilbert's philosophy of mathematics, and the substantially new aspect of his foundational thought from [1922b] onward, was the so-called finitary standpoint. This methodological standpoint consists in a restriction of mathematical thought to objects which are "intuitively present as immediate experience prior to all thought," and to those operations on and methods of reasoning about such objects which do not require the introduction of abstract concepts, in particular, require no appeal to completed infinite totalities.

Hilbert characterized the domain of finitary reasoning in a well-known paragraph which appears in roughly the same formulation in all of Hilbert's more philosophical papers from the 1920s [1922b; 1926; 1928; 1931b]:

> [A]s a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation, certain extra-logical concrete objects that are intuitively present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that can neither be reduced to anything else nor requires reduction. This is the basic philosophical position that I consider req-

> uisite for mathematics and, in general, for all scientific thinking, understanding, and communication. [Hilbert, 1926, 376]

These objects are, for Hilbert, the *signs.* For the domain of contentual number theory, the signs in question are sequences of strokes ("numerals") such as

$$|, ||, |||, ||||| \,.$$

The question of how exactly Hilbert understood the numerals is difficult to answer. What is clear in any case is that they are logically primitive, i.e., they are neither concepts (as Frege's numbers are) nor sets. For Hilbert, the important issue is not primarily their metaphysical status (abstract versus concrete in the current sense of these terms), but that they do not enter into logical relations, e.g., they cannot be predicated of anything. In Bernays's most mature presentations of finitism [Bernays, 1930; Hilbert and Bernays, 1939], the objects of finitism are characterized as *formal objects* which are recursively generated by a process of repetition; the stroke symbols are then concrete representations of these formal objects (see [Sieg, 2002]).

Sometimes Hilbert's view is presented as if Hilbert claimed that the numbers are signs on paper. It is important to stress that this is a misrepresentation, that the numerals are not physical objects in the sense that truths of elementary number theory are dependent only on external physical facts or even physical possibilities (e.g., on what sorts of stroke symbols it is possible to write down). Hilbert [1926] made too much of the fact that for all we know, neither the infinitely small nor the infinitely large are actualized in physical space and time, yet he certainly held that the number of strokes in a numeral is unbounded. It is also essential to the conception that the numerals are sequences of one kind of sign, and that they are somehow dependent on being grasped as such a sequence, that they do not exist independently of our intuition of them. Only our seeing or using "||||" as a sequence of 4 strokes as opposed to a sequence of 2 symbols of the form "||" makes "||||" into the numeral that it is. This raises the question of individuation of stroke symbols. An alternative account would have numerals be mental constructions. However, Bernays denied also this, writing that "the objects of intuitive number theory, the number signs, are, according to Hilbert, also not 'created by thought'. But this does not mean that they exist independently of their *intuitive construction*, to use the Kantian term that is quite appropriate here" [Bernays, 1923, 226]. [Kitcher, 1976] proposes the view that, whatever the numerals are, the strokes on paper or the stroke sequences contemplated by the mind merely *represent* the numerals. According to Hilbert and Bernays, the numerals are given in our representation, but they are not merely subjective "mental cartoons" (Kitcher's term).

> If we want [...] the ordinal numbers as definite objects free of all inessential elements, then in each case we have to take the mere schema of the relevant figure of repetition [*Wiederholungsfigur*] as an object; this requires a very high abstraction. We are free, however, to represent these purely formal objects by concrete objects ("number signs"); these

> contain then inessential, arbitrarily added properties, which, however, are also easily grasped as such. [Bernays, 1930, 244]

One version of this view would be to hold that the numerals are *types* of stroke-symbols as represented in intuition. This is the interpretation that Tait [1981] gives. At first glance, this seems to be a viable reading of Hilbert. It takes care of the difficulties that the reading of numerals-as-tokens (both physical and mental) faces, and it gives an account of how numerals can be dependent on their intuitive construction while at the same time not being created by thought. The reasoning that leads Tait to put forward his reading lies in several constraints that Hilbert and Bernays put on the numerals. For instance, [Bernays, 1923, 159] writes that "figures [i.e., numerals] *are* not shapes, they *have* a shape." Thus it is the shape of the numerals, and not the numerals themselves, which is supposed to be independent of place and time, independent of the circumstances of production, independent of inessential differences in execution, and capable of secure recognition in all circumstances [Hilbert, 1922b, 163]. Tait infers from this that identity between numerals is type identity, and hence, that numerals should be construed as types of stroke symbols.

Types are ordinarily considered to be abstract objects and not located in space or time. Taking the numerals as intuitive representations of sign types might commit us to taking these abstract objects as existing independently of their intuitive representation. That numerals are "space- and timeless" is a consequence that already [Müller, 1923, 158] thought could be drawn from Hilbert's statements, and that was in turn denied by Bernays. The reason is that a view on which numerals are space- and timeless objects existing independently of us would be committed to them existing simultaneously as a completed totality, and this is exactly what Hilbert is objecting to.

> It is by no means compatible, however, with Hilbert's basic thoughts to introduce the numbers as ideal objects "with quite different determinations from those of sensible objects," "which exist entirely independent of us." By this we would go beyond the domain of the immediately certain. In particular, this would be evident in the fact that we would consequently have to assume the numbers *as all existing simultaneously.* But this would mean to assume at the outset that which Hilbert considers to be problematic. [Bernays, 1923, 225–26]

This is not to say that it is *incoherent* to consider the numbers as being abstract objects, only that the finitary viewpoint prohibits such a view. Bernays goes on to say:

> Hilbert's theory does not exclude the possibility of a philosophical attitude which conceives of the numbers [but not the finitist's numerals] as existing, non-sensible objects (and thus the same kind of ideal existence would then have to be attributed to transfinite numbers as well, and in particular to the numbers of the so-called second number class).

> Nevertheless the aim of Hilbert's theory is to make such an attitude dispensable for the foundation of the exact sciences. [Bernays, 1923, 226]

Another open question in this regard is exactly what Hilbert meant by "concrete." He very likely did not use the term in the same sense as it is used today, i.e., as characteristic of spatio-temporal physical objects in contrast to "abstract" objects. However, sign types certainly are different from full-fledged abstracta like pure sets in that all their tokens are concrete. Parsons takes account of this difference by using the term "quasi-concrete" for such abstracta. Tait, on the other hand, thinks that even the tokens are not concrete physical objects, but abstract themselves.

Now what is the epistemological status of the finitary objects? In order to carry out the task of providing a secure foundation for infinitary mathematics, access to finitary objects must be immediate and certain. Hilbert's philosophical background was broadly Kantian, as was Bernays's, who was closely affiliated with the neo-Kantian school of philosophy around Leonard Nelson in Göttingen. Hilbert's characterization of finitism often refers to Kantian intuition, and the objects of finitism as objects given intuitively. Indeed, in Kant's epistemology, immediacy is a defining characteristic of intuitive knowledge. The question is, what kind of intuition is at play? Mancosu [1998b] identifies a shift in this regard. He argues that whereas the intuition involved in Hilbert's early papers was a kind of perceptual intuition, in later writings (e.g., [Bernays, 1928a]) it is identified as a form of pure intuition in the Kantian sense. Hilbert [1931b, 266–267] later sees the finite mode of thought as a separate source of *a priori* knowledge in addition to pure intuition (e.g., of space) and reason, claiming that he has "recognized and characterized the third source of knowledge that accompanies experience and logic." Both Bernays and Hilbert justify finitary knowledge in broadly Kantian terms (without however going so far as to provide a transcendental deduction), characterizing finitary reasoning as the kind of reasoning that underlies all mathematical, and indeed, scientific, thinking, and without which such thought would be impossible.[9]

The simplest finitary propositions are those about equality and inequality of numerals. The finite standpoint moreover allows operations on finitary objects. Here the most basic is that of concatenation. The concatenation of the numerals || and ||| is communicated as "2 + 3," and the statement that || concatenated with ||| results in the same numeral as ||| concatenated with || by "2 + 3 = 3 + 2." In actual proof-theoretic practice, as well as explicitly in [Bernays, 1930; Hilbert and Bernays, 1934], these basic operations are generalized to operations defined by recursion, paradigmatically, primitive recursion, e.g., multiplication and exponentiation. Roughly, a primitive recursive definition of a numerical operation is one in which the function to be defined, f, is given by two equations

$$\begin{aligned} f(0, \mathfrak{m}) &= g(\mathfrak{m}) \\ f(\mathfrak{n}', \mathfrak{m}) &= h(\mathfrak{n}, \mathfrak{m}, f(\mathfrak{n}, \mathfrak{m})), \end{aligned}$$

---

[9] See [Kitcher, 1976] and [Parsons, 1998] for more discussion on the metaphysics and epistemology of finitism, and [Sieg, 1999] for historical remarks on the development of finitism.

where g and h are functions already defined, and $\mathfrak{n}'$ is the successor numeral to $\mathfrak{n}$. Fraktur letters are used here, as in Hilbert's writings, as meta-variables for numerals. For instance, if we accept the functions $g(\mathfrak{m}) = \mathfrak{m}$ (the constant function) and $h(\mathfrak{n}, \mathfrak{m}, \mathfrak{k}) = \mathfrak{m} + \mathfrak{k}$ as finitary, then the equations above define a finitary function, in this case, multiplication $f(\mathfrak{n}, \mathfrak{m}) = \mathfrak{n} \times \mathfrak{m}$.

Similarly, finitary judgments may involve not just equality or inequality but also basic decidable properties, such as "is a prime." This is finitarily acceptable as long as the characteristic function of such a property is itself finitary: For instance, the operation which transforms a numeral to | if it is prime and to || otherwise can be defined by primitive recursion and is hence finitary. Such finitary propositions may be combined by the usual logical operations of conjunction, disjunction, negation, but also bounded quantification. Hilbert [1926] gives the example of the proposition that "there is a prime number between $\mathfrak{p} + 1$ and $\mathfrak{p}! + 1$" where $\mathfrak{p}$ is a certain large prime. This statement is finitarily acceptable since it "serves merely to abbreviate the proposition" that either $\mathfrak{p} + 1$ or $\mathfrak{p} + 2$ or $\mathfrak{p} + 3$ or ... or $\mathfrak{p}! + 1$ is a prime.

The problematic finitary propositions are those that express general facts about numerals such as that $1 + \mathfrak{n} = \mathfrak{n} + 1$ for any given numeral $\mathfrak{n}$. It is problematic because, as Hilbert puts it, it "is from the finitist point of view *incapable of being negated*" [Hilbert, 1926, 378]. By this he means that the contradictory proposition that there is a numeral $\mathfrak{n}$ for which $1 + \mathfrak{n} \neq \mathfrak{n} + 1$ is not finitarily meaningful. "One cannot, after all, try out all numbers" [Hilbert, 1928, 470]. For the same reason, a finitary general proposition is not to be understood as an infinite conjunction but "only as a hypothetical judgment that comes to assert something when a numeral is given" [Hilbert, 1926, 378]. Even though they are problematic in this sense, general finitary statements are of particular importance to Hilbert's proof theory, since the statement of consistency of a formal system T is of such a general form: for any given sequence p of formulas, p is not a derivation of a contradiction in T.

Even though in general existential statements are not finitarily meaningful, they may be *given* finitary meaning if the witness is given by a finitary function. For instance, the finitary content of Euclid's theorem that for every prime $\mathfrak{p}$ there is a prime $> \mathfrak{p}$, is that given a specific prime $\mathfrak{p}$ one can produce, by a finitary operation, another prime $> \mathfrak{p}$ (viz., by testing all numbers between $\mathfrak{p}$ and $\mathfrak{p}! + 1$. This view is discussed by Bernays [1930] and plays an important role in the uses Gentzen [1936] and others make of proof theory.

3.2 Analyses of finitism

Hilbert's substantial philosophical claims about the finitary standpoint are difficult to flesh out. For instance, Hilbert and Bernays both appeal to the role of Kantian intuition for our apprehension of finitary objects (they are given in the faculty of representation). Supposing one accepts this line of epistemic justification in principle, it is plausible that the simplest examples of finitary objects and propositions, and perhaps even simple cases of finitary operations such as concatenations of numerals can be given a satisfactory account. However, it is unclear exactly which more complex objects, propositions, and operations should be admitted as finitary, and how the account can be extended to cover them. This has led to substantial debate since the 1920s about the nature of finitary reasoning and its justification.

Of crucial importance to both an understanding of finitism and of Hilbert's proof theory is the question of what operations and what principles of proof should be allowed from the finitist standpoint. That a general answer is necessary is clear from the demands of Hilbert's proof theory, i.e., it is not to be expected that given a formal system of mathematics (or even a single sequence of formulas) one can "see" that it is consistent (or that it cannot be a genuine derivation of an inconsistency) the way we can see, e.g., that $|| + ||| = ||| + ||$. What is required for a consistency proof is an operation which, given a formal derivation, transforms such a derivation into one of a special form, plus proofs that the operation in fact succeeds in every case and that proofs of the special kind cannot be proofs of an inconsistency. To count as a finitary consistency proof, the operation itself must be acceptable from the finitist standpoint, and the proofs required must use only finitarily acceptable principles.

Hilbert never gave a general account of which operations and methods of proof are acceptable from the finitist standpoint, but only examples of operations and methods of inference in contentual finitary number theory which he accepted as finitary. Contentual induction was accepted in its application to finitary statements of the hypothetical, general kind explicitly in [1922b]. Hilbert [1923, 1139] said that intuitive thought "includes recursion and intuitive induction for finite existing totalities," and used exponentiation in an example in 1928. Bernays [1930] explained how exponentiation may be understood as a finitary operation on numerals. Hilbert and Bernays [1934] give the only general account of finitary contentual number theory; according to it, operations defined by primitive recursion and proofs using induction are finitarily acceptable. All of these methods, in their application in the domain of numbers, can be formalized in a system known as primitive recursive arithmetic (PRA), which allows definitions of functions by primitive recursion and induction on quantifier-free formulas. However, neither Hilbert nor Bernays ever claimed that *only* primitive recursive operations count as finitary, and non-primitive recursive methods were used in ostensibly finitary consistency proofs already in 1923 (see [Tait, 2002] and [Zach, 2003b]). These include, in particular, the consistency proof of a formal system of number theory corresponding to primitive recursive arithmetic in [Hilbert and Bernays, 1923], and a stronger system in Ackermann's dissertation [Ackermann, 1925].[10]

Although Hilbert and his collaborators used methods which go beyond the primitive recursive and accepted them as finitary, it is still unclear whether they (a) realized that these methods were not primitive recursive and (b) whether they would still have accepted them as finitary if they had.[11] The conceptual issue is which operations *should* be considered as finitary. Since Hilbert was less than completely clear on what the finitary standpoint consists in, there is some leeway in setting up the constraints, epistemological and otherwise, an analysis of finitist operation and proof must fulfill. Hilbert characterized the objects of finitary number theory as "intuitively given," as "surveyable in all their parts," and said that their having basic properties must "exist intuitively" for us. Bernays [1922, 216] suggests that in finitary mathematics, only "primitive intuitive cognitions come into play," and uses the term "point of view of intuitive evidence" in connection with finitism [Bernays, 1930, 250]. This characterization of finitism as primarily to do with *intuition* and intuitive knowledge has been emphasized in particular by Parsons [1998] who argues that what can count as finitary on this understanding is not more than those arithmetical operations that can be defined from addition and multiplication using bounded recursion. In particular, according to Parsons, exponentiation and general primitive recursion are not finitarily acceptable.

---

[10] Ackermann in fact used transfinite induction up to $\omega^{\omega^{\omega}}$ to justify a non-primitive recursive reduction procedure.

The thesis that finitism coincides with primitive recursive reasoning has received a forceful and widely accepted defense by Tait [1981]. Tait, in contrast to Parsons, rejects the aspect of representability in intuition as the hallmark of the finitary; instead he takes finitary reasoning to be "a minimal kind of reasoning supposed by all nontrivial mathematical reasoning about numbers" and analyzes finitary operations and methods of proof as those that are implicit in the very notion of number as the form of a finite sequence. This analysis of finitism is supported by Hilbert's contention that finitary reasoning is a precondition for logical and mathematical, indeed, any scientific thinking [Hilbert, 1931b, 267]. The crucial difference between Tait's conception of finitism and Parsons (as well as Hilbert's own) is that according to Tait there is no ultimate epistemological foundation for finitism which yields the security of finitary reasoning for which Hilbert appealed to intuition. Tait argues that

> . . . no absolute conception of security is realized by finitism or any other kind of mathematical reasoning. Rather, the special role of finitism consists in the circumstance that it is a minimal kind of reasoning presupposed by all nontrivial mathematical reasoning about numbers. And for this reason it is *indubitable* in a Cartesian sense that there is no preferred or even equally preferable ground on which to stand and criticize it. Thus finitism is fundamental to number-theoretical mathematics even if it is not a foundation in the sense Hilbert wished. [Tait, 1981, 525]

Another interesting analysis of finitary proof, which, however, does not provide as detailed a philosophical justification, was proposed by Kreisel [1960]. It yields the result that exactly those functions are finitary which can be proved to be well-defined in first-order arithmetic *PA*.[12] Kreisel's proposal makes use of the notions of formalizations of provability predicates and ordinal progressions of theories. Kreisel argues that if $Pr(\ulcorner A \urcorner)$ has been recognized to be a provability predicate for a partial formalization Σ_μ of finitary reasoning, and $Pr(\ulcorner A(0^{(x)}) \urcorner)$[13] is provable in Σ_μ (and hence established by finitary means), then the finitist is entitled to also accept $A(x)$ as finitarily established. If that is the case, we may add $A(x)$ as an axiom to Σ_μ to obtain a new theory Σ_ν which is also finitarily justified. On Kreisel's account, finitary provability coincides with the provability in some Σ_ν so obtained, starting from $\Sigma_0 = PRA$. If some Σ_ν proves $\exists y\, A(x, y)$, for $A(x, y)$ a primitive recursive equation, then $f(x) =$ the least y such that $A(x, y)$, is finitarily justified. Kreisel sketches a proof of the theorem that the functions so justified are exactly those which are provably total in *PA*, and hence there are finitary functions which are not primitive recursive.

---

[11] See Tait's discussion in the Appendix to Chapters 1 and 2 in [Tait, 2005b].

[12] Kreisel [1970, Section 3.5] provides another analysis by focusing on what is "visualizable."

Tait [1981, §13] also contains a discussion of Kreisel's analysis. In order to obtain a non-primitive recursive function on Kreisel's account, we must properly extend Σ_0 since the provably total functions of $\Sigma_0 = PRA$ are just the primitive recursive functions. So suppose we have that *PRA* proves the arithmetization of the claim that $\exists y\, A(0^{(x)}, y)$ is provable.[14] This, according to Kreisel, justifies the acceptance of $f(x)$ as defined above as finitary, because a finitary proof of the general fact that it is provable in *PRA* that $f(x)$ is defined, together with the acceptance of *PRA* as finitarily acceptable, amounts to a finitary proof that $f(x)$ is defined for all x. The weak point in this justification, according to Tait, is this:

> For the finitist to recognize the validity of primitive recursive arithmetic, he must recognize the *general* validity of definition of functions by primitive recursion. But he cannot even formulate this since it involves the notion of function.

Tait's point here is that there is a significant difference between accepting each primitive recursive definition individually as finitary, and accepting primitive recursion in general as a finitarily valid principle. The finitist is able to do the former, but not the latter. For to accept primitive recursion in general as a finitarily valid principle of definition, one would either, as Tait puts it, need to invoke the notion of a function (which is not a finitary object), or one needs a justification for why, say, all primitive recursive *terms* are calculable for every argument — and for this a finitary evaluation procedure for primitive recursive terms is necessary. Such an evaluation procedure, however, cannot be primitive recursive. And prior to the extension of *PRA* to include the new non-primitive recursive function $f(x)$ there is no reason to suppose that such an evaluation procedure exists. Although Tait's objection is directed at Kreisel's analysis of finitary function, it of course also raises doubts about Kreisel's account of finitary proof.

The result is the same: finitary functions turn out to be just those provably total in *PA*.

[13] Here, x is a free variable, and $\ulcorner A(0^{(x)}) \urcorner$ is the function of x which computes $\ulcorner A(0'^{\cdots\prime}) \urcorner$ with x occurrences of $'$.

[14] In other words, there are primitive recursive functions $h(x)$ and $g(x)$ so that *PRA* proves that $g(x)$ codes a derivation in *PRA* of the formula $\ulcorner A(0^{(x)}, t) \urcorner$, where t is the primitive recursive term (containing only the free variable x) which is coded by $h(x)$.

3.3 Hilbert's program and instrumentalism

Hilbert's program has often been interpreted as an instrumentalist account of mathematics. This reading relies on the distinction Hilbert makes between the finitary part of mathematics and the non-finitary rest which is in need of grounding (via finitary meta-mathematics). The finitary part Hilbert calls "contentual," i.e., its propositions and proofs have content. The infinitary part, on the other hand, is "not meaningful from a finitary point of view." This distinction corresponds to a distinction between formulas of the the axiomatic systems of mathematics for which consistency proofs are being sought. Some of the formulas correspond to contentual, finitary propositions: they are the "real" formulas. The rest are called "ideal." They are added to the real part of our mathematical theories in order to preserve classical inferences such as the principle of the excluded middle for infinite totalities, i.e., the principle that either all numbers have a given property or there is a number which does not have it. This disjunction is not finitarily valid, as we saw above. Hilbert first mentioned "ideal" propositions in [1926], although the distinction was hinted at in [1923]. In the latter paper, Hilbert presented a formal system of quantifier-free number theory about which he says that "the provable formulae we acquire in this way all have the character of the finite" (1139). Then the transfinite axioms (i.e., quantifiers) are added to simplify and complete the theory (1144). Here he draws the analogy with the method of ideal elements: "In my proof theory, the transfinite axioms and formulae are adjoined to the finite axioms, just as in the theory of complex variables the imaginary elements are adjoined to the real, and just as in geometry the ideal constructions are adjoined to the actual" (ibid). When Hilbert, in [1926], explicitly introduces the notion of an ideal proposition, and in [1928], when he first speaks of *real propositions* in addition to the ideal, he is quite clear that the real part of the theory consists only of decidable, variable-free formulas. They are supposed to be "directly capable of verification" — akin to propositions derived from laws of nature which can be checked by experiment [Hilbert, 1928, 475].[15] It is this extension of the real part of the theory by the ideal, infinitary part that is in need of justification by a consistency proof:

> For there is a condition, a single but absolutely necessary one, to which the use of the method of ideal elements is subject, and that is the *proof of consistency*; for, extension by the addition of ideals is legitimate only if no contradiction is thereby brought about in the old, narrower domain, that is, if the relations that result for the old objects whenever the ideal objects are eliminated are valid in the old domain. [Hilbert, 1926, 383]

Weyl [1925] described Hilbert's project as replacing meaningful mathematics by a meaningless game of formulas. He noted that Hilbert wanted to "secure not *truth*, but the *consistency* of analysis" and suggested a criticism that echoes an earlier one by Frege: Why should we take consistency of a formal system of mathematics as a reason to believe in the truth of the pre-formal mathematics it codifies? Is Hilbert's meaningless inventory of formulas not just "the bloodless ghost of analysis"? Weyl suggested a solution:

---

[15] This reading is not universally accepted. [Detlefsen, 1990], for instance, considers the real formulas to also include the general formulas, i.e., formulas with free variables. See [Zach, 2004] for a defense of the reading given here.

> [I]f mathematics is to remain a serious cultural concern, then some *sense* must be attached to Hilbert's game of formulae, and I see only one possibility of attributing to it (including its transfinite components) an independent intellectual meaning. In theoretical physics we have before us the great example of a [kind of] knowledge of completely different character than the common or phenomenal knowledge that expresses purely what is given in intuition. While in this case every judgment has its own sense that is completely realizable within intuition, this is by no means the case for the statements of theoretical physics. In that case it is rather *the system as a whole* that is in question if confronted with experience. [Weyl, 1925, 140]

The analogy with physics is striking. Hilbert himself used a similar analogy in [1928]. He there suggested that consistency is not the only virtue ideal mathematics has: transfinite inference simplifies and abbreviates proofs, "brevity and economy of thought are the *raison d'être* of existence proofs" (476). The formal system of transfinite logic is not arbitrary: "This formula game is carried out according to certain definite rules, in which the *technique of our thinking* is expressed... The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds" (ibid).

Although these remarks are suggestive, they do not force an interpretation of Hilbert as an instrumentalist. Most commentators have taken this reading (including [**?**; Giaquinto, 1983; Sieg, 1990], and in particular [Detlefsen, 1986]), in that they interpret Hilbert's explanation that the ideal propositions "have no meaning in themselves" [Hilbert, 1926, 381] as claiming that classical mathematics is a *mere* instrument, and that statements of transfinite mathematics have no truth value. To the extent that this is accurate, however, it must be understood as a methodological instrumentalism: A successful execution of the proof-theoretic program would show that one could pretend *as if* mathematics was meaningless.[16] The analogy with physics is therefore not: transfinite propositions have no meaning just as propositions involving theoretical terms have no meaning, but: transfinite propositions require no direct intuitive meaning just as one does not have to directly see electrons in order to theorize about them. Hallett [1990], taking into account the 19th century mathematical background from which Hilbert came as well as published and unpublished sources from Hilbert's entire career (in particular [Hilbert, 1992], the most extensive discussion of the method of ideal elements) comes to the following conclusion:

---

[16] On this point see also [Sieg, 1999], esp. B3 and the conclusion, and [Sieg, 2002].

> [Hilbert's treatment of philosophical questions] is *not* meant as a kind of instrumentalist agnosticism about existence and truth and so forth. On the contrary, it is meant to provide a non-skeptical and positive solution to such problems, a solution couched in cognitively accessible terms. And, it appears, the same solution holds for both mathematical and physical theories. Once new concepts or "ideal elements" or new theoretical terms have been accepted, then they exist in the sense in which *any* theoretical entities exist. [Hallett, 1990, 239]

This conclusion is in line with Weyl's assessment in [1928]. When Weyl eventually turned away from intuitionism,[17] he emphasized the purpose of Hilbert's proof theory, not to turn mathematics into a meaningless game of symbols, but to turn it into a theoretical science which codifies scientific (mathematical) practice.

The reading of Hilbert as an instrumentalist goes hand in hand with a reading of the proof-theoretic program as a reductionist project. The instrumentalist reading interprets ideal mathematics as a meaningless formalism, which simplifies and "rounds out" mathematical reasoning. But a consistency proof of ideal mathematics by itself does not explain what ideal mathematics is an instrument *for*. Thus, commentators have sought to elaborate on Hilbert's discussion of consistency by pointing out that consistency proofs do not just establish that ideal theories are free from formal contradictions, but that they establish more than mere consistency. They establish *conservativity* of the ideal over the real part of the theory, in the following sense: If the ideal theory proves a real statement, then the real statement is also provable by real, finitary means. Such reductivist projects were common in the philosophy of science at the time, as was pointed out by Giaquinto [1983]. A conservativity proof justifies the use of transfinite mathematics: it is not only internally consistent, but it proves only true intuitive propositions.

On this picture, classical mathematics is to be formalized in a system which includes formalizations of all the directly verifiable (by calculation) propositions of contentual finite number theory. The consistency proof should show that all real propositions which can be proved by ideal methods are true, i.e., can be directly verified by finite calculation. Actual proofs such as the ε-substitution procedure are of such a kind: they provide finitary procedures which eliminate transfinite elements from proofs of real statements. In particular, they turn putative ideal derivations of $0 = 1$ into derivations in the real part of the theory; the impossibility of such a derivation establishes consistency of the theory. Indeed, Hilbert saw that something stronger is true: not only does a consistency proof establish truth of real formulas provable by ideal methods, but it yields finitary proofs of finitary *general* propositions if the corresponding free-variable formula is derivable by ideal methods [Hilbert, 1928, 474].

It bears pointing out that Hilbert never clearly articulated conservativity of the ideal over the real for finitary general statements as an aim of his foundational project. There are contemporary commentators, e.g., von Neumann [1931] who attribute to Hilbert an interest in conservativity proofs, but it was only Bernays in the *Grundlagen der Mathematik* who pointed out that consistency proofs themselves established not only the truth of variable-free formulas provable by ideal methods, but also of free-variable theorems. In this context, Bernays used the term 'verifiable' (*verifizierbar*): a free-variable formula $A(x)$ is verifiable if each numerical instance $A(\mathfrak{z})$ is (finitarily) true. The proof transformation methods used in consistency proofs for systems of arithmetic in [Hilbert and Bernays, 1934, 248, 298] can be applied not only to putative proofs of $0 = 1$, but generally to proofs of formulas with free-variables. If we have a proof of $A(x)$, then the following method constitutes a finitary proof that, for any $\mathfrak{z}$, $A(\mathfrak{z})$ is true. From the derivation of $A(x)$ we obtain a derivation of $A(\mathfrak{z})$ by substitution. The procedure given in the consistency proof transforms this derivation into a variable-free derivation of $A(\mathfrak{z})$ in the real part of the theory, which codifies a finitary calculation that $A(\mathfrak{z})$ is true.

---

[17] For the reasons for Weyl's rejection of intuitionism, see [Mancosu and Ryckman, 2002].

Kreisel was most influential in promoting the interpretation of the aim of Hilbert's program as an attempt to establish conservativity of the ideal theory for finitary general propositions all along. Kreisel [1951] cites Bernays's results; but in [Kreisel, 1960] and later, he instead points to the argument in [Hilbert, 1928, 474]. This argument, unlike Bernays's, does not rely on a particular form of the consistency proof. It assumes only that a finitary consistency proof for an ideal theory is available. Assume there is a derivation of $A(x)$. Now suppose that for a given $\mathfrak{z}$, $A(\mathfrak{z})$ is not true. Then $\neg A(\mathfrak{z})$ would be true,[18] and so there would be a derivation of $\neg A(\mathfrak{z})$ in the ideal theory (which includes all real theorems). But from the derivation of $A(x)$ we obtain, by substitution, a derivation of $A(\mathfrak{z})$, and hence a contradiction. Since we assume that we have a finitary consistency proof of T, this cannot be the case. Hence, $A(\mathfrak{z})$ must be true.

3.4 Hilbert's program and Gödel's incompleteness theorems

Gödel announced the second incompleteness theorem in an abstract published in October 1930: no consistency proof of systems such as *Principia*, Zermelo-Fraenkel set theory, or the systems investigated by Ackermann and von Neumann is possible by methods which can be formulated in these systems. In the full version of his paper, Gödel [1931] left open the possibility that there could be finitary methods which are not formalizable in these systems and which would yield the required consistency proofs. Bernays's first reaction in a letter to Gödel in January 1931 was likewise that "if, as von Neumann does, one takes it as certain that any and every finitary consideration may be formalized within the system P — like you, I regard that in no way as settled — one comes to the conclusion that a finitary demonstration of the consistency of P is impossible" [Gödel, 2003a, 87].

Through a careful ("Gödel"-) coding of sequences of symbols (formulas, proofs) by numbers, Gödel showed that in theories T which contain a sufficient amount of arithmetic, it is possible to produce a formula $Pr(x, y)$ which expresses that x is (the code of) a proof of (the formula with code) y. Specifically, if $\ulcorner 0 = 1 \urcorner$ is the code of the formula $0 = 1$, then $Con_T \equiv \forall x \, \neg Pr(x, \ulcorner 0 = 1 \urcorner)$ expresses that T is consistent (no number is the code of a derivation in T of $0 = 1$). The second incompleteness theorem (G2) says that under certain assumptions about T and the coding apparatus, T does not prove Con_T. Now suppose there were a finitary consistency proof of T. The methods used in such a proof would presumably be formalizable in T. ("Formalizable" means that, roughly, if the proof uses a finitary operation f on derivations which transforms any derivation d into a derivation $f(d)$ of a simple form; then there is a formula $F(x, y)$ so that, for all derivations d, $T \vdash F(\ulcorner d \urcorner, \ulcorner f(d) \urcorner)$.) The consistency of T would be finitarily expressed as the general hypothetical claim that, if d is any given sequence of symbols, d is not a derivation in T of the formula $0 = 1$. The formalization of this proposition is the formula $\neg Pr(x, \ulcorner 0 = 1 \urcorner)$ in which the variable x occurs free. If there were a finitary proof of the consistency of T, its formalization would yield a derivation in T of $\neg Pr(x, \ulcorner 0 = 1 \urcorner)$, from which Con_T can be derived in T by simple universal generalization on x. Yet, a derivation of Con_T in T is ruled out by G2.

---

[18] This inference uses tertium non datur, but only regarding the unproblematic finitary statement $A(\mathfrak{z})$.

Gödel and Bernays initially thought that the difficulty for the consistency proof of Peano arithmetic *PA* could be overcome by employing methods which, although not formalizable in *PA*, are nonetheless finitary. Bernays did not seem to have any particular candidates for such a method in mind, and he thought that all methods which were up to then employed in finitary considerations were in fact formalizable in *PA*. Another option he considered was an extension of the notion of an axiomatic theory by a finitary version of the ω-rule proposed by Hilbert [1931a; 1931b]. Such theories might avoid the impact of Gödel's incompleteness theorem since they do not satisfy the conditions of the incompleteness theorems: the set of axioms would not be decidable. It is fair to say, however, that since about 1934 it has been almost universally accepted that the methods of proof accepted as finitary prior to Gödel's results are all formalizable in *PA* and that the incompleteness theorems do show that there can be no finitary consistency proofs for axiomatic theories of the kind originally considered by Hilbert.

The reaction to the incompleteness theorems in the Hilbert school then focused on extensions of the original finitary standpoint in which consistency proofs for substantial theories like *PA* can be carried out. Such extensions have been proposed and defended on broadly finitary grounds, e.g., Gentzen [1936] defended the use of transfinite induction up to ε_0 in his consistency proof for *PA* as "indisputable," and Takeuti [1987] gave another defense. In the Gentzen-Schütte tradition of proof theory by ordinal analysis, the proof methods used to give consistency proofs are all of this sort. To wit, one uses transfinite induction on ordinal notation systems which name larger and larger ordinals. The more complicated the ordering, the more difficult it is to see that the induction principle in question is finitarily justified. Another extension of the finitary standpoint is due to Gödel [1958].

Smoryński [1977], following earlier suggestions by Kreisel, has argued that already the first incompleteness theorem defeats Hilbert's program. This argument uses the fact that a finitary consistency proof of an ideal theory T implies the conservativity of T over finitary, real mathematics for general finitary statements of the form $A(x)$ (with free variable x). Now Gödel's first incompleteness theorem (G1) states that for any sufficiently strong, consistent formal theory S there is a sentence G_S which is not derivable in S if S is consistent. G_S is a general real sentence. Consider a theory T which formalizes ideal mathematics and contains the theory S, which formalizes real mathematics, as a subtheory. S satisfies the conditions of G1 and hence S does not derive G_S. Yet T, being a formalization of all of mathematics, proves (via a formalization of G1) that if S is consistent, then G_S, but it also proves the consistency (indeed, the soundness) of S. Hence T proves G_S. Thus, we have a true real statement which is provable in ideal mathematics but not in real mathematics.[19]

4 HILBERT'S PROGRAM NOW

4.1 Detlefsen's Hilbertian instrumentalism

Detlefsen [1979; 1986; 2001] has proposed a wide-ranging instrumentalist account of mathematics based on Hilbert's program which is designed to escape the difficulties that Gödel's incompleteness theorems poses for Hilbert's original consistency project. The project has several parts: Detlefsen [1986] first gives a detailed analysis and defense of a general instrumentalist view of mathematics along Hilbertian lines. This includes an analysis of the epistemic import of ideal proofs of real statements, which answers a question that was hardly addressed by Hilbert, either in his mature writings in the 1920s or in his exchange with Frege on formalism and consistency. This is the question of how manipulation of meaningless strings of symbols can ever lead to knowledge (of finitary truths). Detlefsen then analyzes the characteristics of formal systems of ideal mathematics *qua* instruments. On Detlefsen's view, even though, say, full set theory is commonly accepted as a formalization of infinitary mathematics, only parts of set theory are in fact instrumentally useful. In particular, (1) ideal proofs of real theorems which are more complex than any real proof of the same theorem do not yield an instrumental advantage, and hence are not instrumentally useful; and (2) ideal proofs which are too long or complex to be comprehended by humans, and hence never play a role in actual mathematical reasoning, are also of no instrumental value. A proof theoretic justification of instrumental mathematics, i.e., the proof of the conservativity of the ideal theory over real mathematics, is only required, so Detlefsen, for the instrumentally useful part. Detlefsen introduces the term "Hilbertian residue" for that part of ideal mathematics that is instrumentally useful and hence in need of proof-theoretic justification. On his view, then, we only need a consistency proof for the Hilbertian residue, not for all of ideal mathematics.

---

[19] The argument appeals to a number of perhaps contentious assumptions, such as that T proves the soundness of S. For a critique, see [Detlefsen, 1990].

This move from a required justification for all of infinitary mathematics to a justification of only the Hilbertian residue is one step toward Detlefsen's defense of instrumentalism against Gödel's incompleteness theorems. For the incompleteness theorems only apply under certain conditions, e.g., only when the theory in question contains enough basic arithmetic to carry out Gödel coding, formalization of a proof predicate, and to prove the diagonal lemma. The Hilbertian residue of a theory, however, need not contain a sufficiently strong arithmetical subtheory because of (1) above. This provides part of Detlefsen's defense against the challenge posed by the first incompleteness theorem [Detlefsen, 1986, Appendix]. Detlefsen [1990] also argues that instrumentalism escapes the argument from G1 by denying that ideal mathematics must be conservative over the real part. According to him, Hilbertian instrumentalism requires only that the ideal theory not prove anything which is in conflict with the real theory; it is not required that all its real theorems are also provable by real means. If this defense is successful, Detlefsen is right to claim that not G1, but only G2 poses a real challenge to instrumentalism.

Detlefsen presents several lines of defense against the argument from G2, one of which [1979] echoes [Hilbert, 1931b]. If a version of the ω-rule is finitarily acceptable, then we would have found a finitarily acceptable method of proof which is not capable of formalization. Hence, real mathematics is not a subtheory of the ideal instrument, but this was an assumption necessary to draw the conclusion that there can be no real consistency proof of the ideal theory. Ignjatovič [1994] raised serious doubts about the acceptability of Detlefsen's version of the ω-rule, however. Detlefsen's other argument against the common interpretation of Gödel's second theorem focuses on the notion of formalization. That the particular formalization of "T is consistent" by Gödel's formula Con_T is not provable does not imply that there could not be other formulas, which *are* provable in T, and which have as much right to be called "formalizations of the consistency of T." These rely on different formalizations of the provability predicate Pr_T than the standard ones. It is known that formalized consistency statements are unprovable whenever the provability predicate obeys certain general derivability conditions. Detlefsen argues that these conditions are not necessary for a predicate to count as a genuine provability predicate, and indeed there are provability predicates which violate the provability conditions and which give rise to consistency formulas which *are* provable in their corresponding theories. These, however, depend on nonstandard conceptions of provability which would likely not have been accepted by Hilbert. One quite basic example is the use of Rosser provability instead of ordinary provability. On this approach, a derivation of a formula A only counts as a proof if no derivation with smaller Gödel number is a derivation of $\neg A$. If $Prov(x, \ulcorner A \urcorner)$ is the standard formalization of "x is the code of a derivation of the formula A," then the Rosser provability predicate is given by

$$RPr(\ulcorner A \urcorner) \equiv \exists x(Prov(x, \ulcorner A \urcorner) \land \forall y < x \, \neg Prov(y, \ulcorner \neg A \urcorner)).$$

For this provability predicate, $\neg RPr(\ulcorner 0 = 1 \urcorner)$ *is* provable in, e.g., first-order Peano arithmetic. Provability of a formula A, however, is no longer just a matter of deriving it from the axioms; one also has to check that all shorter derivations do not end in $\neg A$. Other "consistency minded" theories which prove their own consistency are discussed, e.g., in [Jeroslow, 1971; Jeroslow, 1975] and especially [Visser, 1989]. The Rosser provability predicate is studied in, e.g., [Guaspari and Solovay, 1979] and [Arai, 1990].[20]

Another interesting aspect of Detlefsen's approach to instrumentalism and Hilbert's program related to technical work in proof theory is the emphasis on instrumental utility of ideal proofs. Hilbert, as we saw above, himself noted the theoretical and cognitive advantage of ideal methods, such as increased simplicity of proofs. In Detlefsen's instrumentalism, such considerations take center stage. Even if it is conceded that Gödel's theorems call the success of instrumentalism in its most general form into question, it would still be of substantial interest to study restricted cases of conservative extensions of real mathematics which are instrumentally useful. To flesh out the notion of "instrumental usefulness," one obvious characteristic of formal proofs is length. For instance, one might take an ideal theory to be useful if its proofs are substantially shorter than proofs in, say, *PRA* of the same theorems. This question is amenable to precise proof theoretical study. Caldon and Ignjatovič [2005] prove some related, but on the whole, negative results: The subsystem of first-order arithmetic $I\Sigma_1$ in which induction is limited to Σ_1 formulas has super-exponential "speed-up" over *PRA*. This indicates that using induction over non-finitary formulas (Σ_1 formulas have unbounded existential quantifiers) yields significantly shorter proofs than proofs without. However, more comprehensive theories (*RCA*, *WKL*, see below) which contain some second-order apparatus, do not significantly shorten proofs vis-à-vis $I\Sigma_1$.

4.2 Generalized Hilbert programs

The work of Gentzen on consistency proofs for mathematical theories using methods that go beyond the strictly finitary marks the beginning of an important line of proof-theoretic research. As outlined in 2.3 above, Gentzen's approach was to retain the aim of Hilbert's program, viz., to give consistency proofs for strong mathematical theories by restricted means. Because of Gödel's incompleteness theorems, these restricted means are necessarily not themselves formalizable in the theories whose consistency is established by them. Nevertheless, they retain a constructive character, and attempts have been made to justify them on finitary grounds.

The consistency proof of Gentzen [1936], as discussed above, uses the principle of transfinite induction up to ε_0 in order to establish the consistency of first-order Peano arithmetic. Gentzen's use of a system of notations for ordinals less than ε_0, and the proof of the termination of a reduction procedure for derivations in *PA* based on induction on these ordinal notations, provide the model for the proof theoretic analysis of axiomatic systems along these lines. In order to give an "ordinal analysis" of a theory T, one typically produces an ordinal notation system for ordinals less than some ordinal α such that for every $\beta < \alpha$, the formalization $TI(\beta)$ of the transfinite induction principle for β is provable in T. In practice, using transfinite induction up to α itself and otherwise only strictly finitary methods, one can prove the consistency of T. The fact that induction up to ε_0 establishes the consistency of *PA*, together with the result of [Gentzen, 1943] that shows that for all $\beta < \varepsilon_0$, *PA* proves $TI(\beta)$ for all $\beta < \varepsilon_0$ constitutes an ordinal analysis of *PA*, and we say that ε_0 is the proof theoretic ordinal of *PA*.

---

[20] For technical background, discussion of intensional provability predicates and examples, see [Feferman, 1960]. For discussion, see also [Resnik, 1974b], [Auerbach, 1985; Auerbach, 1992] and [Steiner, 1991].

Proof theory in the tradition of Gentzen and Schütte as well as Takeuti has focused on such ordinal analyses of theories of increasing strength. In recent work, Rathjen [2005b; 2005a] has pushed the boundaries of this approach in giving an ordinal analysis of a very strong subsystem of analysis called Π^1_2-comprehension.[21] The consistency proofs in this tradition are, for the most part, based on the approach of [Schütte, 1960], which uses a variant of Gentzen's formalization using infinitary derivations. A second tradition has pursued ordinal analysis using extensions of Ackermann's ε-substitution method [1940], for examples see [Mints and Tupailo, 1999] and [Arai, 2003].

Although generalized Hilbert programs in this tradition have certainly produced important mathematical work, its philosophical underpinnings are thin. Takeuti [1987] attempted to give a finitary justification for the proof theoretic ordinal ε_0, but it is unclear to what extent more complex ordinal notation systems are finitarily acceptable. Even if one concedes, as, e.g., Schütte does, that the consistency proofs in question are constructive (although no longer strictly finitary), it is still unclear what the philosophical significance of the technical results is. Feferman [1988, 366] offers this assessment:

> [A]s the systems of ordinal notation used for consistency proofs of stronger and stronger theories become more and more complicated, the significance to noncognoscenti of what is thereby accomplished decreases in inverse proportion. Thus, on the one hand, to say that one has obtained a constructive consistency proof of a theory T — without saying anything more — is too general to be informative; and, on the other hand, to say that the proof has been carried out by transfinite induction on a certain complicated recursive ordering for some very large ordinal tells us nothing about what constructive principles are involved in the proof of this well-ordering.[22]

Another important proof-theoretical approach in which the analysis of systems of classical mathematics is accomplished using a generalization of the finitary standpoint is that of functional interpretations. The model for this approach is Gödel's *Dialectica* interpretation [1958]. The *Dialectica* interpretation shows how one can reduce an infinitary theory T (in this case, intuitionistic first-order arithmetic) to a quantifier-free theory F.[23] An ordinal analysis of a theory does something similar, for instance, one can view Gentzen's consistency proof as reducing Peano arithmetic to a quantifier-free theory (PRA) extended by a certain infinitary induction principle ($TI(\varepsilon_0)$). In the case of functional interpretations, the quantifier-free theory F is also not strictly finitary: it does not just mention finite objects but also certain infinitary objects, viz., functionals of finite type. A functional interpretation can be seen as a reduction of the infinitary theory T to the theory of functionals F in question. The approach using functional interpretations has the following advantage over the Gentzen-Schütte approach. It is a consequence of the reduction of T to F that every recursive function which can be proved to be total in T is represented by a term of F. Because the functionals of F in practice characterize natural classes of functions, a functional interpretation yields an appealing analysis of the computational content of F. Moreover, the conceptual import of the reduction is more apparent than in the case of ordinal analysis: already in the case of PA, Gödel's functionals of finite type provide a much clearer account of the character of the constructive methods appealed to than induction up to ε_0.[24]

---

[21] See [Pohlers, 1987] for a survey of the work in the Schütte school, and [Pohlers, 1998] for a more recent technical survey.

[22] For a more forceful criticism of proof theory in this tradition, see [Kreisel, 1976].

4.3 Relativized Hilbert programs

A philosophically more satisfactory continuation of Hilbert's program in proof theoretic terms has been suggested by Kreisel [1954; 1968; 1983] and has been elaborated especially by Feferman. This work proceeds from a wider conception of Hilbert's program as an attempt to justify ideal mathematics by restricted means. On this conception, the aim of Hilbert's proof theory was to show that, at least as far as a certain class of real propositions is concerned, ideal mathematics does not go beyond real mathematics, and in this sense finitary mathematics is a foundation for ideal mathematics. A finitary consistency proof of the kind envisaged by Hilbert would have accomplished this for all of classical mathematics.

The scope of the foundational project, however, need not necessarily be all of higher mathematics. So-called relativized Hilbert programs are projects in which one considers certain fragments of higher mathematics as the theory for which a foundation is sought (and, indeed, also theories stronger than finitism as candidates for the reducing theory, e.g., predicative theories). Examples of these are Feferman's work on explicit mathematics and predicative subsystems of analysis, and to some extent also the Friedman-Simpson program of reverse mathematics (see below). What is common to these approaches to mathematical foundations is that they concentrate on so-called proof-theoretic reductions of systems of classical mathematics to more restricted systems. The reduction is carried out using finitist means, and therein lies its philosophical significance.

---

[23] Via the interpretation of classical arithmetic in intuitionistic arithmetic [Gentzen, 1933; Gödel, 1933], the *Dialectica* interpretation also yields a functional interpretation of classical arithmetic.

[24] For an excellent survey of this approach, see [Avigad and Feferman, 1998].

A foundational reduction in Feferman's sense [1988; 1993a] is accomplished if it can be shown that a body of mathematics which is justified by a foundational framework $\mathcal{F}_1$ (e.g, finitary, constructive, predicative, infinitary, set-theoretic) can already be justified, in a certain sense, in a weaker, or stricter foundational framework $\mathcal{F}_2$. This is in general not possible in a wholesale fashion, however, partial foundational reductions can and have been achieved. Suppose a theory T_1 is justified by a foundational framework $\mathcal{F}_1$, and a theory T_2 by a weaker framework $\mathcal{F}_2$. A proof theoretic reduction of T_1 to T_2 (conservative for Φ) is a partial recursive function f such that

1. whenever x is (the code of) a proof in T_1 of a formula (with code) y in Φ, then $f(x)$ is (the code of) a proof of y in T_2, and
2. T_2 proves the formalization of (1).

If there is such a function f, we write $T_1 \leq T_2[\Phi]$. Now if T_1 is directly justified by a foundational framework $\mathcal{F}_1$, and T_2 by $\mathcal{F}_2$, then, so Feferman, a proof-theoretic reduction that establishes $T_1 \leq T_2[\Phi]$ is a partial foundational reduction of $\mathcal{F}_1$ to $\mathcal{F}_2$. Clause (2) in the definition ensures that the reduction (the function f) itself is justified by the weaker framework $\mathcal{F}_2$. In the reductions achieved in practice, it turns out that f is actually primitive recursive and the formalization of (1) can even be proved in primitive recursive arithmetic *PRA*. Since *PRA* is directly justified by the finitary framework, such partial foundational reductions are therefore all finitarily justified. Feferman's main philosophical conclusion from the possibility of giving such foundational reductions is this: The main argument for set-theoretical realism is the Quine-Putnam indispensability argument, which proceeds from the premises that set-theory is indispensable to science. Feferman has shown, first, that much, if not all, of scientifically applicable mathematics can actually be formalized in much weaker systems (e.g., Feferman's system W, which is justified by a predicative foundational framework), and second, that predicative mathematics can be reduced to the countably infinite (in the sense that there is a partial foundational reduction of predicative mathematics to countably infinite mathematics, given by a proof-theoretic reduction of W to Peano Arithmetic *PA*). He concludes that,

> even if one accepts the indispensability argument, practically nothing philosophically definite can be said of the entities which are then supposed to have the same status — ontologically and epistemologically — as the entities of natural science. That being the case, what do the indispensability arguments amount to? As far as I'm concerned, they are completely vitiated. [Feferman, 1993b]

Independently of the question of mathematical realism and of the scope and force of the indispensability arguments, proof-theoretic reductions give precise answers to questions of the relation between foundational frameworks. Since a proof-theoretic reduction of T_1 to T_2 also yields a consistency proof of T_1 in T_2 (i.e., a relative consistency result), establishing a proof-theoretic reduction also provides a solution to Hilbert's program relativized to T_1 and T_2. Feferman summarizes the importance of proof-theoretic reductions thus:

> In general, the kinds of results presented here serve to sharpen what is to be said in favor of, or in opposition to, the various philosophies of mathematics such as finitism, predicativism, constructivism, and set-theoretical realism. Whether or not one takes one or another of these philosophies seriously for ontological and/or epistemological reasons, it is important to know which parts of mathematics are in the end justifiable on the basis of the respective philosophies and which are not. The uninformed common view — that adopting one of the non-platonistic positions means pretty much giving up mathematics as we know it — needs to be drastically corrected, and that should also no longer serve as the last-ditch stand of set-theoretical realism. On the other hand, would-be nonplatonists must recognize the now clearly marked sacrifices required by such a commitment and should have well-thought out reasons for making them. [Feferman, 1993a]

Proof theorists have obtained a number of such results, including reductions of theories which on their face require a significant amount of ideal mathematics for their justification (e.g., subsystems of analysis) to finitary systems.[25]

The program of so-called reverse mathematics developed by, in particular, Friedman and Simpson, is another continuation of Hilbert's program. In the face of Gödel's results showing that not all of classical mathematics can be reduced to the finitary, they seek to answer the question: how much of classical mathematics can be so reduced? Reverse mathematics aims to give a precise answer to this question by investigating which theorems of classical mathematics are provable in weak subsystems of analysis which are reducible to finitary mathematics (in the sense discussed above). A typical result is that the Hahn-Banach theorem of functional analysis is provable in a theory known as WKL_0 (for "weak König lemma"); WKL_0 is proof-theoretically reducible to PRA for Π^0_2 sentences (i.e., sentences of the form $\forall x \exists y\, A(x, y)$).[26]

Reverse mathematics in this tradition is primarily concerned with infinitary theories, e.g., subsystems of analysis. Gödel's theorems show, however, that not even all truths of first-order number theory are provable in Peano arithmetic, and hence that not even the domain of all arithmetical truths can be given a foundation on finitary principles. This suggests the questions of whether there are "mathematically interesting" statements of number theory which are not provable in systems that can be reduced to the finitary. The most celebrated result in this regard is the proof by Paris and Harrington [1977] that a version of the finite Ramsey theorem is not provable in Peano arithmetic. However, this and other examples of independent number theoretic statements are constructed specifically to be independent of Peano arithmetic. It turns out that a great many "ordinary" number theoretic results are provable even in weak fragments of first-order number theory, and this has led Friedman to conjecture that "every theorem published in the *Annals of Mathematics* whose statement involves only finitary mathematical objects (i.e., what logicians call an arithmetical statement) can be proved in elementary arithmetic." (Here, elementary arithmetic is a very weak theory which can be proved consistent by primitive recursive methods.) Avigad [2003] gives an excellent survey of the issues and results related to this conjecture and places it in the philosophical context of Hilbert's program.[27]

---

[25] For a discussion of the philosophical significance of such proof theoretic reductions, see [Feferman, 2000] and [Hofweber, 2000].

[26] See [Simpson, 1988] for an overview, [Simpson, 1999] for a technical introduction to reverse mathematics, and also the collection [Simpson, 2005].

The results surveyed in this section are all natural continuations of Hilbert's original ideas. A central aspect of recent proof-theoretical investigations and of Hilbert's original program alike is that they study formalized systems using metamathematical tools with the aim of understanding the structure and content of these systems. Hilbert's original consistency project, the conservativity project that Kreisel and others interpret Hilbert as having been engaged in, as well as reductive proof theory are all examples of this, and this is also part of the reason why many practicing proof theorists see themselves as still working on Hilbert's program. Ordinal analysis, functional interpretations, proof theoretic reductions and reverse mathematics are only some of the most prominent areas of proof theory, and those most explicitly situated in the tradition of Hilbert's program. Many other areas of proof theory other than those directly concerned with consistency and foundational reductions of theories are related to the aims of Hilbert's program, e.g., the no-counterexample interpretation [Kreisel, 1951; Tait, 2005a] and work on the structure and complexity of formal proofs more generally [Pudlák, 1998].

5 CONCLUSION

Although it has been a commonplace in the literature on the philosophy of mathematics and logic from 1950 onward that Hilbert's program has not only been "killed" by Gödel's incompleteness theorems but that it was over-ambitions if not ill-conceived from the start, in the current literature a more positive evaluation has emerged. This is in large part due to the attention which unpublished writings in the Hilbert school (especially lecture notes to Hilbert's courses) have received recently, as well as to the availability of more of the published writings in English translation (e.g., in [Ewald, 1996] and [Mancosu, 1998a]). But it is also due to a growing recognition that the common characterizations of Hilbert's program are caricatures, and to a clearer philosophical engagement with the recent results of proof theory. For instance, Ramsey's characterization that, according to Hilbert, "Mathematics proper is thus regarded as a sort of game, played with meaningless marks on paper rather like noughts and crosses" [Ramsey, 1926, 231], and the view that Hilbert held a naive formalist and instrumentalist view of mathematics, have been criticized by various writers. It remains to be seen in what ways Hilbert's philosophical views can be resuscitated (in the manner in which, e.g., Frege's logicist program has experienced a renaissance in the neo-logicist writings of, e.g., Boolos, Heck, and Hale and Wright). It should be clear in any case from the discussion in the preceding section that ideas closely related to Hilbert's own *have* been hugely successful. And it is also clear from the recent historical studies of Hilbert's unpublished writings as well as from the study of the proof theoretical practice of the Hilbert school that the ideas which underpin much of recent and current proof theoretical research are not merely "inspired by" Hilbert's program. Hilbert's fundamental aim was, all along, to make mathematical reasoning amenable to mathematical investigation, and to carry out such an investigation which yields an analysis of non-constructive reasoning in terms of restricted methods. Hilbert, of course, emphasized consistency of non-constructive systems as the most interesting property to be investigated, and emphasized finitary methods as those in terms of which such an analysis should be carried out. But even in the 1920s, in the practice of consistency proofs in the work of Ackermann, Bernays, and von Neumann, among others, more broadly constructive methods were employed in this analysis, and results about properties other than consistency were obtained. Gentzen's work of the 1930s and subsequent proof theoretical studies should thus not be seen as merely a response to Gödel's incompleteness results, but more broadly as advancing Hilbert's original aim of investigating the structure of mathematical reasoning. Seen in this light, again, proof theory as a foundational enterprise is very much alive. Although Gödel's theorems show that Hilbert's original expectations about what exactly can be analyzed in which way and with what restricted methods can not be fulfilled, proof theory and Hilbert's aims more generally have been advanced tremendously over the last half-century.

---

[27] See also [Raatikainen, 2003] on the current status of the various branches of proof-theoretic research relating to Hilbert's program.

ACKNOWLEDGEMENTS

This paper is a substantially revised and expanded version of my entry in the *Stanford Encyclopedia of Philosophy* [Zach, 2003a]. I would like to acknowledge the support of the Social Sciences and Humanities Research Council of Canada, as well as helpful comments from Jeremy Avigad, Aleksandar Ignjatovič, Paolo Mancosu, and William Tait.

BIBLIOGRAPHY

[Ackermann, 1925] W. Ackermann. Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit. *Mathematische Annalen*, 93:1–36, 1925.

[Ackermann, 1940] W. Ackermann. Zur Widerspruchsfreiheit der Zahlentheorie. *Mathematische Annalen*, 117:162–194, 1940.

[Arai, 1990] T. Arai. Derivability conditions on Rosser's provability predicates. *Notre Dame Journal of Formal Logic*, 31:487–497, 1990.

[Arai, 2003] T. Arai. Epsilon substitution method for $ID_1(\Pi^0_1 \vee \Sigma^0_1)$. *Annals of Pure and Applied Logic*, 121:163–208, 2003.

[Auerbach, 1985] D. Auerbach. Intentionality and the Gödel theorems. *Philosophical Studies*, 48:337–351, 1985.

[Auerbach, 1992] D. Auerbach. How to say things with formalisms. In Michael Detlefsen, editor, *Proof, Logic, and Formalization*, pages 77–93. Routledge, London, 1992.

[Avigad, 2003] J. Avigad. Number theory and elementary arithmetic. *Philosophia Mathematica*, 11:257–284, 2003.

[Avigad and Feferman, 1998] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation. In Buss [1998].

[Avigad and Reck, 2001] J. Avigad and E. Reck. "Clarifying the nature of the infinite": the development of metamathematics and proof theory. Technical report cmu-phil-120, Carnegie Mellon University, 2001. Online at `http://www.andrew.cmu.edu/user/avigad/Papers/infinite.pdf`.

[Avigad and Zach, 2002] J. Avigad and R. Zach. The epsilon calculus. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Summer 2002. `http://plato.stanford.edu/entries/epsilon-calculus/`.

[Barwise, 1977] J. Barwise, editor. *Handbook of Mathematical Logic.* North-Holland, Amsterdam, 1977.

[Benacerraf and Putnam, 1983] P. Benacerraf and H. Putnam, editors. *Philosophy of Mathematics.* Cambridge University Press, Cambridge, 2nd edition, 1983.

[Bernays, 1922] P. Bernays. Über Hilberts Gedanken zur Grundlegung der Arithmetik. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 31:10–19, 1922. English translation in [Mancosu, 1998a, 215–222].

[Bernays, 1923] P. Bernays. Erwiderung auf die Note von Herrn Aloys Müller: Über Zahlen als Zeichen. *Mathematische Annalen*, 90:159–63, 1923. English translation in [Mancosu, 1998a, 223–226].

[Bernays, 1928a] P. Bernays. Über Nelsons Stellungnahme in der Philosophie der Mathematik. *Die Naturwissenschaften*, 16:142–45, 1928.

[Bernays, 1928b] P. Bernays. Zusatz zu Hilberts Vortrag über "Die Grundlagen der Mathematik". *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 6:88–92, 1928. English translation in: [van Heijenoort, 1967, 485–489].

[Bernays, 1930] P. Bernays. Die Philosophie der Mathematik und die Hilbertsche Beweistheorie. *Blätter für deutsche Philosophie*, 4:326–67, 1930. Reprinted in [Bernays, 1976, 17–61]. English translation in [Mancosu, 1998a, 234–265].

[Bernays, 1970] P. Bernays. On the original Gentzen consistency proof. In A. Kino, J. Myhill, and R. E. Veseley, editors, *Intuitionism and Proof Theory*, pages 409–417. North-Holland, Amsterdam, 1970.

[Bernays, 1976] P. Bernays. *Abhandlungen zur Philosophie der Mathematik.* Wissenschaftliche Buchgesellschaft, Darmstadt, 1976.

[Buss, 1998] S. R. Buss, editor. *Handbook of Proof Theory.* Elsevier, Amsterdam, 1998.

[Caldon and Ignjatovič, 2005] P. Caldon and A. Ignjatovič. On mathematical instrumentalism. *Journal of Symbolic Logic*, 70:778–794, 2005.

[Detlefsen, 1979] M. Detlefsen. On interpreting Gödel's second theorem. *Journal of Philosophical Logic*, 8:297–313, 1979. Reprinted with a postscript in [Shanker, 1988, 131–154].

[Detlefsen, 1986] M. Detlefsen. *Hilbert's Program.* Reidel, Dordrecht, 1986.

[Detlefsen, 1990] M. Detlefsen. On an alleged refutation of Hilbert's program using Gödel's first incompleteness theorem. *Journal of Philosophial Logic*, 19:343–377, 1990.

[Detlefsen, 2001] M. Detlefsen. What does Gödel's second theorem say? *Philosophia Mathematica*, 9:37–71, 2001.

[Detlefsen, 2005] M. Detlefsen. Formalism. In Stewart Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 236–317. Oxford University Press, New York and Oxford, 2005.

[Ewald, 1996] W. B. Ewald, editor. *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, volume 2. Oxford University Press, Oxford, 1996.

[Feferman, 1960] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, pages 35–92, 1960.

[Feferman, 1988] S. Feferman. Hilbert's Program relativized: Proof-theoretical and foundational reductions. *Journal of Symbolic Logic*, 53(2):364–284, 1988.

[Feferman, 1993a] S. Feferman. What rests on what? The proof-theoretic analysis of mathematics. In Johannes Czermak, editor, *Philosophy of Mathematics. Proceedings of the Fifteenth International Wittgenstein-Symposium, Part 1*, pages 147–171, Vienna, 1993. Hölder-Pichler-Tempsky. Reprinted in [Feferman, 1998, Ch. 10, 187–208].

[Feferman, 1993b] S. Feferman. Why a little bit goes a long way: Logical foundations of scientifically applicable mathematics. *Proceedings of the Philosophy of Science Association 1992*, 2:442–455, 1993. Reprinted in [Feferman, 1998, Ch. 14, 284–298].

[Feferman, 1998] S. Feferman. *In the Light of Logic*. Oxford University Press, New York and Oxford, 1998.

[Feferman, 2000] Solomon Feferman. Does reductive proof theory have a viable rationale? *Erkenntnis*, 53:63–96, 2000.

[Field, 1980] H. Field. *Science without Numbers. A Defence of Nominalism*. Princeton University Press, Princeton, 1980.

[Gentzen, 1933] G. Gentzen. Über das Verhältnis zwischen intuitionistischer und klassischer Logik. First published in *Archiv für mathematische Logik und Grundlagenforschung* 16 (1974), 119–132. English translation in [Gentzen, 1969, 53–67], 1933.

[Gentzen, 1934] G. Gentzen. Untersuchungen über das logische Schließen I–II. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1934. English translation in [Gentzen, 1969, 68–131].

[Gentzen, 1935] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. Published as "Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie" in *Archiv für mathematische Logik und Grundlagenforschung* 16 (1974), 97–118. English translation in [Gentzen, 1969, 132–213], 1935.

[Gentzen, 1936] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 112:493–565, 1936. English translation in [Gentzen, 1969, 132–213].

[Gentzen, 1943] G. Gentzen. Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, 119:140–161, 1943. English translation in [Gentzen, 1969, 287–308].

[Gentzen, 1969] G. Gentzen. *The Collected Papers of Gerhard Gentzen*, M. E. Szabo, editor. North-Holland, Amsterdam, 1969.

[Giaquinto, 1983] M. Giaquinto. Hilbert's philosophy of mathematics. *British Journal for Philosophy of Science*, 34:119–132, 1983.

[Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. Reprinted and translated in [Gödel, 1986, 144–195].

[Gödel, 1933] K. Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematisches Kolloquiums*, 4:34–38, 1933. Reprinted and translated in [Gödel, 1986, 286–295].

[Gödel, 1958] K. Gödel. Über eine bisher noch nicht benütze Erweiterung des finiten Standpunktes. *Dialectica*, pages 280–287, 1958. Reprinted and translated in [Gödel, 1995, 217–251].

[Gödel, 1986] K. Gödel. *Collected Works*, volume 1, Solomon Feferman et al., editors. Oxford University Press, Oxford, 1986.

[Gödel, 1995] K. Gödel. *Collected Works*, volume 2, Solomon Feferman et al., editors. Oxford University Press, Oxford, 1995.

[Gödel, 2003a] K. Gödel. *Collected Works*, volume 4, Solomon Feferman et al., editors. Oxford University Press, Oxford, 2003.

[Gödel, 2003b] K. Gödel. *Collected Works*, volume 5, Solomon Feferman et al., editors. Oxford University Press, Oxford, 2003.

[Guaspari and Solovay, 1979] D. Guaspari and R. M. Solovay. Rosser sentences. *Annals of Mathematical Logic*, 16:81–99, 1979.

[Hallett, 1990] M. Hallett. Physicalism, reductionism and Hilbert. In Andrew D. Irvine, editor, *Physicalism in Mathemtics*, pages 183–257. Reidel, Dordrecht, 1990.

[Hallett, 1994] M. Hallett. Hilbert's axiomatic method and the laws of thought. In Alexander George, editor, *Mathematics and Mind*, pages 158–200. Oxford University Press, Oxford, 1994.

[Herbrand, 1930] J. Herbrand. *Recherches sur la théorie de la démonstration.* Doctoral dissertation, University of Paris, 1930. English translation in [Herbrand, 1971, 44–202].

[Herbrand, 1971] J. Herbrand. *Logical Writings*, Warren D. Goldfarb, editor. Harvard University Press, 1971.

[Hilbert, 1899] D. Hilbert. Grundlagen der Geometrie. In *Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen*, pages 1–92. Teubner, Leipzig, 1st edition, 1899. reprinted in [Hilbert, 2004].

[Hilbert, 1900a] D. Hilbert. Mathematische Probleme. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Klasse*, pages 253–297, 1900. Lecture given at the International Congress of Mathematicians, Paris, 1900. Partial English translation in [Ewald, 1996, 1096–1105].

[Hilbert, 1900b] D. Hilbert. Über den Zahlbegriff. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 8:180–84, 1900. English translation in [Ewald, 1996, 1089–1096].

[Hilbert, 1905] D. Hilbert. Über die Grundlagen der Logik und der Arithmetik. In A. Krazer, editor, *Verhandlungen des dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, pages 174–85, Leipzig, 1905. Teubner. English translation in [van Heijenoort, 1967, 129–38].

[Hilbert, 1918a] D. Hilbert. Axiomatisches Denken. *Mathematische Annalen*, 78:405–15, 1918. Lecture given at the Swiss Society of Mathematicians, 11 September 1917. Reprinted in [Hilbert, 1935, 146–56]. English translation in [Ewald, 1996, 1105–1115].

[Hilbert, 1918b] D. Hilbert. Prinzipien der Mathematik. Lecture notes by Paul Bernays. Winter-Semester 1917–18. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen. Published in [Hilbert, 2006], 1918.

[Hilbert, 1920] D. Hilbert. Probleme der mathematischen Logik. Vorlesung, Sommer-Semester 1920. Lecture notes by Paul Bernays and Moses Schönfinkel. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen. Published in [Hilbert, 2006], 1920.

[Hilbert, 1922a] D. Hilbert. Grundlagen der Mathematik. Vorlesung, Winter-Semester 1921–22. Lecture notes by Paul Bernays. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen. Published in [Hilbert, 2006], 1922.

[Hilbert, 1922b] D. Hilbert. Neubegründung der Mathematik: Erste Mitteilung. *Abhandlungen aus dem Seminar der Hamburgischen Universität*, 1:157–77, 1922. Series of talks given at the University of Hamburg, July 25–27, 1921. Reprinted with notes by Bernays in [Hilbert, 1935, 157–177]. English translation in [Mancosu, 1998a, 198–214] and [Ewald, 1996, 1115–1134].

[Hilbert, 1923] D. Hilbert. Die logischen Grundlagen der Mathematik. *Mathematische Annalen*, 88:151–165, 1923. Lecture given at the Deutsche Naturforscher-Gesellschaft, September 1922. Reprinted in [Hilbert, 1935, 178–191]. English translation in [Ewald, 1996, 1134–1148].

[Hilbert, 1926] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95:161–90, 1926. Lecture given Münster, 4 June 1925. English translation in [van Heijenoort, 1967, 367–392].

[Hilbert, 1928] D. Hilbert. Die Grundlagen der Mathematik. *Abhandlungen aus dem Seminar der Hamburgischen Universität*, 6:65–85, 1928. English translation in [van Heijenoort, 1967, 464-479].

[Hilbert, 1929] D. Hilbert. Probleme der Grundlegung der Mathematik. *Mathematische Annalen*, 102:1–9, 1929. Lecture given at the International Congress of Mathematicians, 3 September 1928. English translation in [Mancosu, 1998a, 227–233].

[Hilbert, 1931a] D. Hilbert. Beweis des Tertium non datur. *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen. Math.-phys. Klasse*, pages 120–25, 1931.

[Hilbert, 1931b] D. Hilbert. Die Grundlegung der elementaren Zahlenlehre. *Mathematische Annalen*, 104:485–94, 1931. Reprinted in [Hilbert, 1935, 192–195]. English translation in [Ewald, 1996, 1148–1157].

[Hilbert, 1935] D. Hilbert. *Gesammelte Abhandlungen*, volume 3. Springer, Berlin, 1935.

[Hilbert, 1992] D. Hilbert. *Natur und mathematisches Erkennen.* Birkhäuser, Basel, 1992. Vorlesungen, 1919–20.

[Hilbert, 2004] D. Hilbert. *David Hilbert's Lectures on the Foundations of Geometry, 1891–1902*, Ulrich Majer and Michael Hallett, editors. Springer, New York, 2004.

[Hilbert, 2006] D. Hilbert. *David Hilbert's Lectures on the Foundations of Arithmetic and Logic, 1917–1933*, William Ewald and Wilfried Sieg, editors. Springer, New York, 2006. forthcoming.

[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik.* Springer, Berlin, 1928.

[Hilbert and Bernays, 1923] D. Hilbert and P. Bernays. Logische Grundlagen der Mathematik. Vorlesung, Winter-Semester 1922–23. Lecture notes by Paul Bernays, with handwritten notes by Hilbert. Hilbert-Nachlaß, Niedersächsische Staats- und Universitätsbibliothek, Cod. Ms. Hilbert 567. Published in [Hilbert, 2006], 1923.

[Hilbert and Bernays, 1934] D. Hilbert and P. Bernays. *Grundlagen der Mathematik*, volume 1. Springer, Berlin, 1934.

[Hilbert and Bernays, 1939] D. Hilbert and P. Bernays. *Grundlagen der Mathematik*, volume 2. Springer, Berlin, 1939.

[Hofweber, 2000] T. Hofweber. Proof-theoretic reduction as a philosopher's tool. *Erkenntnis*, 53:127–146, 2000.

[Ignjatovič, 1994] A. Ignjatovič. Hilbert's program and the omega-rule. *Journal of Symbolic Logic*, 59:322–343, 1994.

[Jeroslow, 1971] R. G. Jeroslow. Consistency statements in formal theories. *Fundamenta Mathematicae*, 72:17–40, 1971.

[Jeroslow, 1975] R. G. Jeroslow. Experimental logics and Δ^0_2-theories. *Journal of Philosophical Logic*, 4:253–267, 1975.

[Kitcher, 1976] P. Kitcher. Hilbert's epistemology. *Philosophy of Science*, 43:99–115, 1976.

[Kreisel, 1951] G. Kreisel. On the interpretation of non-finitist proofs. Part I. *Journal of Symbolic Logic*, 16:241–267, 1951.

[Kreisel, 1954] G. Kreisel. A variant to Hilbert's theory of the foundations of arithmetic. *British Journal for the Philosophy of Science*, 4:107–129, 1954.

[Kreisel, 1960] G. Kreisel. Ordinal logics and the characterization of informal notions of proof. In J. A. Todd, editor, *Proceedings of the International Congress of Mathematicians. Edinburgh, 14–21 August 1958*, pages 289–299, Cambridge, 1960. Cambridge University Press.

[Kreisel, 1968] G. Kreisel. A survey of proof theory. *Journal of Symbolic Logic*, 33:321–388, 1968.

[Kreisel, 1970] G. Kreisel. Principles of proof and ordinals implicit in given concepts. In A. Kino, J. Myhill, and R. E. Veseley, editors, *Intuitionism and Proof Theory*, pages 489–516. North-Holland, Amsterdam, 1970.

[Kreisel, 1971] G. Kreisel. Review of *The Collected Papers of Gerhard Gentzen* [Gentzen, 1969]. *Journal of Philosophy*, 68:238–265, 1971.

[Kreisel, 1976] G. Kreisel. Wie die Beweistheorie zu ihren Ordinalzahlen kam und kommt. *Jahresberichte der Deutschen Mathematiker-Vereinigung*, 78:177–223, 1976.

[Kreisel, 1983] G. Kreisel. Hilbert's programme. In Benacerraf and Putnam [1983], pages 207–238.

[Mancosu and Ryckman, 2002] P. Mancosu and T. Ryckman. Mathematics and phenomenology: The correspondence between O. Becker and H. Weyl. *Philosophia Mathematica*, 10:130–202, 2002.

[Mancosu, 1998a] P. Mancosu, editor. *From Brouwer to Hilbert. The Debate on the Foundations of Mathematics in the 1920s.* Oxford University Press, Oxford, 1998.

[Mancosu, 1998b] P. Mancosu. Hilbert and Bernays on metamathematics. In *From Brouwer to Hilbert* [1998a], pages 149–188.

[Mancosu, 1999] P. Mancosu. Between Russell and Hilbert: Behmann on the foundations of mathematics. *Bulletin of Symbolic Logic*, 5(3):303–330, 1999.

[Mancosu, 2003] P. Mancosu. The Russellian influence on Hilbert and his school. *Synthèse*, 137(1–2):59–101, 2003.

[Mancosu, 2004] P. Mancosu. Review of Kurt Gödel, *Collected Works*, vols. IV and V, Solomon Feferman, et al., eds. Oxford: Oxford University Press, 2003. *Notre Dame Journal of Formal Logic*, 45:109–125, 2004.

[Mints and Tupailo, 1999] G. Mints and S. Tupailo. Epsilon-substitution method for the ramified language and Δ^1_1-comprehension rule. In Andrea Cantini et al., editors, *Logic and Foundations of Mathematics*, pages 107–130. Kluwer, Dordrecht, 1999.

[Moore, 1997] G. H. Moore. Hilbert and the emergence of modern mathematical logic. *Theoria (Segunda Época)*, 12:65–90, 1997.

[Müller, 1923] A. Müller. Über Zahlen als Zeichen. *Mathematische Annalen*, 90:153–158; 163, 1923.

[Negri, 1980] M. Negri. Constructive sequent reduction in Gentzen's first consistency proof for arithmetic. In Maria Luisa Dalla Chiara, editor, *Italian Studies in the Philosophy of Science*, pages 153–168. Reidel, Dordrecht, 1980.

[Paris and Harrington, 1977] J. Paris and L. Harrington. A mathematical incompleteness in Peano arithmetic. In Barwise [1977], pages 1133–1142.

[Parsons, 1998] C. Parsons. Finitism and intuitive knowledge. In Matthias Schirn, editor, *The Philosophy of Mathematics Today*, pages 249–270. Oxford University Press, Oxford, 1998.

[Peckhaus, 1990] V. Peckhaus. *Hilbertprogramm und Kritische Philosophie.* Vandenhoeck und Ruprecht, Göttingen, 1990.

[Pohlers, 1987] W. Pohlers. Contributions of the Schütte school school in Munich to proof theory, 1987. Appendix to [Takeuti, 1987].

[Pohlers, 1998] W. Pohlers. Subsystems of set theory and second-order number theory. In Buss [1998].

[Poincaré, 1906] H. Poincaré. Les mathématiques et la logique. *Revue de métaphysique et de morale*, 14:294–317, 1906. English translation in [Ewald, 1996, 1038–1052].

[Pudlák, 1998] P. Pudlák. The length of proofs. In Buss [1998].

[Raatikainen, 2003] Panu Raatikainen. Hilbert's program revisited. *Synthèse*, 137:157–177, 2003.

[Ramsey, 1926] F. P. Ramsey. Mathematical logic. *The Mathematical Gazette*, 13:185–94, 1926. Reprinted in [Ramsey, 1990, 225–244].

[Ramsey, 1990] F. P. Ramsey. *Philosophical Papers*, D. H. Mellor, editor. Cambridge University Press, Cambridge, 1990.

[Rathjen, 2005a] M. Rathjen. An ordinal analysis of parameter free Π^1_2-comprehension. *Archive for Mathematical Logic*, 44:263–362, 2005.

[Rathjen, 2005b] M. Rathjen. An ordinal analysis of stability. *Archive for Mathematical Logic*, 44:1–62, 2005.

[Resnik, 1974a] M. D. Resnik. The Frege-Hilbert controversy. *Philosophy and Phenomenological Research*, 34:386–403, 1974.

[Resnik, 1974b] M. D. Resnik. On the philosophical significance of consistency proofs. *Journal of Philosophical Logic*, 3:133–47, 1974.

[Resnik, 1980] M. D. Resnik. *Frege and the Philosophy of Mathematics.* Cornell University Press, Ithaca, 1980.

[Russell, 1902] B. Russell. Letter to Frege, June 16, 1902. In van Heijenoort [1967], pages 124–125.

[Schütte, 1960] K. Schütte. *Beweistheorie.* Springer, Berlin, 1960.

[Shanker, 1988] S. G. Shanker. *Gödel's Theorem in Focus.* Routledge, London, 1988.

[Sieg, 1990] W. Sieg. Reflections on Hilbert's program. In Wilfried Sieg, editor, *Acting and Reflecting*, pages 171–82. Kluwer, Dordrecht, 1990.

[Sieg, 1999] W. Sieg. Hilbert's programs: 1917–1922. *Bulletin of Symbolic Logic*, 5(1):1–44, 1999.

[Sieg, 2002] W. Sieg. Beyond Hilbert's reach? In David B. Malament, editor, *Reading Natural Philosophy. Essays in the History and Philosophy of Science and Mathematics*, pages 363–405. Open Court, Chicago and La Salle, Ill., 2002.

[Simpson, 1988] S. G. Simpson. Partial realizations of Hilbert's program. *Journal of Symbolic Logic*, 53(2):349–363, 1988.

[Simpson, 1999] S. G. Simpson. *Subsystems of Seond Order Arithmetic.* Springer, Berlin, 1999.

[Simpson, 2005] S. G. Simpson. *Reverse Mathematics 2001.* Lecture Notes in Logic 21. Association for Symbolic Logic and A K Peters, Wellesley, Mass., 2005.

[Smoryński, 1977] C. Smoryński. The incompleteness theorems. In Barwise [1977], pages 821–865.

[Stein, 1988] Howard Stein. *Logos*, logic, and *logisitiké*: Some philosophical remarks on nineteenth-century transformation of mathematics. In William Aspray and Philip Kitcher, editors, *History and Philosophy of Modern Mathematics*, volume 11 of *Minesota Studies in the Philosophy of Science*, pages 238–259. University of Minnesota Press, 1988.

[Steiner, 1975] M. Steiner. *Mathematical Knowledge.* Cornell University Press, Ithaca, 1975.

[Steiner, 1991] M. Steiner. Review of *Hilbert's Program: An Essay on Mathematical Instrumentalism* [Detlefsen, 1986]. *Journal of Philosophy*, 88(6):331–336, 1991.

[Tait, 1981] W. W. Tait. Finitism. *Journal of Philosophy*, 78:524–546, 1981. Reprinted in [Tait, 2005b, 21–42].

[Tait, 2002] W. W. Tait. Remarks on finitism. In Wilfried Sieg, Richard Sommer, and Carolyn Talcott, editors, *Reflections on the Foundations of Mathematics. Essays in Honor of Solomon Feferman*, Lecture Notes in Logic 15. Association for Symbolic Logic and A K Peters, 2002. Reprinted in [Tait, 2005b, 43–53].

[Tait, 2005a] W. W. Tait. Gödel's reformulation of Gentzen's first consistency proof for arithmetic: the no-counterexample interpretation. *Bulletin of Symbolic Logic*, 11:225–238, 2005.

[Tait, 2005b] W. W. Tait. *The Provenance of Pure Reason. Essays in the Philosophy of Mathematics and Its History.* Oxford University Press, New York, 2005.

[Takeuti, 1987] G. Takeuti. *Proof Theory.* Studies in Logic 81. North-Holland, Amsterdam, 2nd edition, 1987.

[van Heijenoort, 1967] J. van Heijenoort, editor. *From Frege to Gödel. A Source Book in Mathematical Logic, 1897–1931.* Harvard University Press, Cambridge, Mass., 1967.

[Visser, 1989] A. Visser. Peano's smart children: A provability logical study of systems with built-in consistency. *Notre Dame Journal of Formal Logic*, 30:161–196, 1989.

[von Neumann, 1927] J. von Neumann. Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift*, 26:1–46, 1927.

[von Neumann, 1931] J. von Neumann. Die formalistische Grundlegung der Mathematik. *Erkenntnis*, 2:116–34, 1931. English translation in: [Benacerraf and Putnam, 1983, 61–65].

[Webb, 1997] J. C. Webb. Hilbert's formalism and arithmetization of mathematics. *Synthèse*, 110:1–14, 1997.

[Weyl, 1921] Hermann Weyl. Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift*, 10:37–79, 1921. Reprinted in [Weyl, 1968, 143-180]. English translation in [Mancosu, 1998a, 86–118].

[Weyl, 1925] H. Weyl. Die heutige Erkenntnislage in der Mathematik. *Symposion*, 1:1–23, 1925. Reprinted in: [Weyl, 1968, 511–42]. English translation in: [Mancosu, 1998a, 123–42].

[Weyl, 1928] H. Weyl. Diskussionsbemerkungen zu dem zweiten Hilbertschen Vortrag über die Grundlagen der Mathematik. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 6:86–88, 1928. English translation in [van Heijenoort, 1967, 480-484].

[Weyl, 1968] H. Weyl. *Gesammelte Abhandlungen*, volume 1, K. Chandrasekharan, editor. Springer Verlag, Berlin, 1968.

[Whitehead and Russell, 1910] A. N. Whitehead and B. Russell. *Principia Mathematica*, volume 1. Cambridge University Press, Cambridge, 1910.

[Whitehead and Russell, 1912] A. N. Whitehead and B. Russell. *Principia Mathematica*, volume 2. Cambridge University Press, Cambridge, 1912.

[Whitehead and Russell, 1913] A. N. Whitehead and B. Russell. *Principia Mathematica*, volume 3. Cambridge University Press, Cambridge, 1913.

[Zach, 1999] R. Zach. Completeness before Post: Bernays, Hilbert, and the development of propositional logic. *Bulletin of Symbolic Logic*, 5(3):331–366, 1999.

[Zach, 2003a] R. Zach. Hilbert's program. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Fall 2003. `http://plato.stanford.edu/archives/entries/hilbert-program/`.

[Zach, 2003b] R. Zach. The practice of finitism. Epsilon calculus and consistency proofs in Hilbert's Program. *Synthèse*, 137:211–259, 2003.

[Zach, 2004] R. Zach. Hilbert's "Verunglückter Beweis," the first epsilon theorem and consistency proofs. *History and Philosophy of Logic*, 25:79–94, 2004.

LOGICISM AND ITS CONTEMPORARY LEGACY

Herbert Hochberg

While logicism, as a philosophy of mathematics, is traditionally taken to begin with Frege, throughout the 20^{th} century it has been associated with Whitehead and Russell's classic *Principia Mathematica* (1910-1913).[1] Ramsey had cited Russell's theory of definite descriptions as a "paradigm" of philosophical analysis, but Russell's logistic analysis of elementary arithmetic was also a significant factor in the development of logical analysis, linguistic precision, and logical reconstruction in 20^{th} century philosophy. It is fitting, then, that, as we shall later consider, the two prominent kinds of "incomplete symbols" in *Principia* are definite descriptions and class abstracts (the latter the key to the *Principia* variant of logicism).[2] Russell's theory of descriptions, Frege's philosophy of language and Alexius Meinong's theories of objects and assumptions were landmark contributions to 20^{th} century discussions and attempts to resolve long standing philosophical problems regarding reference and meaning, the relation of thoughts and judgments to their "objects" and the perennial paradoxes posed by truth and reference. In a parallel way, Russell's logistic analysis of arithmetic, dealing not only with the classical issues in the philosophy of arithmetic — "What are numbers?" and "What kind of truths are arithmetical "truths"?" — but with the puzzles posed by the development of set theory and the explorations of what Hilbert termed "Cantor's paradise," and advocating a method of "logical construction," was another major influence on the emergence of logical positivism and analytic philosophy. *Principia Mathematica* has also assumed a pre-eminent place in the history of logic, along with works of Aristotle and Frege. The two governing ideas of logicism are, first, that the truths of standard arithmetic, properly analyzed, can be understood to be logical truths and, second, that all basic arithmetical terms (concepts) can be "defined" in the vocabulary of (analyzed in terms of) the terms (concepts) of standard logic (a problematic anticipation of second-order predicate logic, in Frege's case, a ramified type variant in *Principia Mathematica*, and the set-theoretical versions in the cumulative work of Zermelo, Bernays, Gödel, von Neumann, Frankel). Assume one considers, as Russell did, the familiar axioms of Peano (1889):[3]

---

[1] The work was originally intended to be a second volume of Russell's 1903 *The Principles of Mathematics*, (*Principia*, 1950, v).

[2] That classes, via Russell's "no-class" theory, were considered "incomplete symbols" has been somewhat ignored by critics of logicism, while the fact that *propositions* form a third kind of "incomplete symbol" has been almost totally overlooked. The failure to recognize these "kinds" as incomplete symbols makes a number of discussions of Russell's supposed commitment to propositional entities in *Principia*, as well as some criticisms of his logicism, pointless. Russell uses the phrase "incomplete symbol" to speak of both signs and what signs supposedly represent — the first use indicates the sign is eliminable, the second that the (kind of) seemingly "represented" entity does not exist.

[3] [Russell, 1956, 125]. Russell uses "finite integer" in place of "number" but then replaces the former by

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

(P-D) 1. 0 is a number.

2. If *x is a number*, the *successor of x is a number*.
3. If the *successor of x* = the *successor of y*, *x* = *y*.
4. If *x is a* number, 0 is not the *successor of x*.
5. If 0 is a ϕ and the successor of every number that is a ϕ is a ϕ, then every *number* is a ϕ.

(P-D) contains three primitive expressions, "0", "*number*" and "*successor of*" that do not belong to the vocabulary of a standard system of logic — as "if. . . then," "and," "the" and "any" do belong to that vocabulary. As (P-D) is taken to provide an adequate *foundation* for the arithmetic of addition and multiplication, either via the familiar "inductive definitions" or by *explicit* definitions in a higher order or set-theoretical framework, the logicist takes the task of showing that arithmetical truths are logical truths to be two-fold. First, "defining" the three primitives solely in terms of logical signs; second, employing the definitions to derive (P-D) as theorems within a standard logical framework. The definitions would serve to provide an "analysis" of the basic concepts of arithmetic, while the derivation of (P-D), so understood, would show that the purported foundational statements of arithmetic are logical truths [Russell, 1903, 124-125]. Russell expressed these "goals" in his preface [1902] to the first edition of *The Principles*:

> The present work has two main objects. One of these, the proof that all pure mathematics deals exclusively with concepts definable in terms of a very small number of fundamental logical concepts, and that all its propositions are deducible from a very small number of fundamental logical principles. . . . .
>
> The other object of this work, which occupies Part I., is the explanation of the fundamental concepts which mathematics accepts as indefinable. This is a purely philosophical task. . . . The discussion of indefinables — which forms the chief part of philosophical logic — is the endeavour to see clearly, and to make others see clearly, the entities concerned, in order that the mind may have that kind of acquaintance with them which it has with redness or the taste of a pineapple. Where, as in the present case, the indefinables are obtained primarily as the necessary residue in a process of analysis, it is often easier to know that there must be such entities than actually to perceive them; there is a process analogous to that which resulted in the discovery of Neptune, with the difference that the final stage — the search with a mental

---

the latter. Though these particular statements are in the form of Peano's postulates, such axioms are now often called the "Peano-Dedekind axioms" due to Dedekind's earlier construction (1887/88) of an alternative system. Frege, discussing Dedekind's book, commented that while Dedekind thought that "the theory of numbers is a part of logic," he did not reduce the concepts he used to "acknowledged logical notions" [Frege, 1967, 4]. What Frege apparently means is that Dedekind did not take cardinal numbers to be classes or any objects characterized apart from his axioms — though he did take the set of (ordinal) numbers to be the greatest common part of all hereditary sets containing the initial element — a variant of the Frege-Russell definition of "natural number" [Kneale, 1962, 470-472].

> telescope for the entity which has been inferred — is often the most difficult part of the undertaking. In the case of classes, I must confess, I have failed to perceive any concept fulfilling the conditions requisite for the notion of *class*. And the contradiction discussed in Chapter x. proves that something is amiss, but what this is I have hitherto failed to discover [1903, xv-xvi].

The contradiction is what first came to be known as *Russell's Paradox* and, more recently, *The Russell-Zermelo Paradox*.[4] Dedekind, by contrast, can be seen as a kind of structuralist who took his axioms to specify, and thereby identify, an infinite domain or "system," as he called it, of elements ordered as a progression or "omega sequence." This raises questions about such elements — as objects arrived at by "abstraction" or "creations of the human mind" or merely imagined possibilities, since Dedekind's focus is on the structure of the system. However they are taken, the elements only have the properties and relations imposed by the axioms for a progression. Such a view has similarities to a view of a contemporary logician, P. Lindström, that we will discuss, and which he calls "quasi-realism." (Lindström, however, briefly comments on and separates his view from Dedekind's [2000, 146].) Later in *The Principles*, Russell, as Frege had, criticized Dedekind's structuralism:

> . . . it is impossible that the ordinals should be, as Dedekind suggests, nothing but the terms of such relations as constitute a progression. If they are anything at all, they must differ from other entities as points from instants or colours from sounds. What Dedekind intended to indicate was probably a definition by means of the principle of abstraction, such as we attempted to give in the preceding chapter. But a definition so made always indicates some class of entities having (or being) a genuine nature of their own, and not logically dependent upon the manner in which they have been defined [1956, 249].

This issue is not over the supposed need for having properties other than relational properties but about being objects of a definite kind that are not simply specified as objects that satisfy the axioms for a progression and are thereby simply distinguished by their place in the ordering. What Russell means by the "principle of abstraction" here is the use of "similarity of two classes" (one-one correlation) to be "analyzable (by the principle of abstraction) into possession of a common property. This we define as the number of either class" [1956, 241]. Thus his rejection of Dedekind's structuralism in *The Principles* stems, in part, from his holding that an equivalence relation (for Russell an instantiated symmetrical and transitive relation) is "constituted by possession of a common property" [1956, 166]. This was aided by the fact that "when I wrote the "Principles," I shared with Frege a belief in the Platonic reality of numbers which, in my imagination, peopled the timeless realm of Being" [1956, x].

---

[4] The paradox stems from Frege's Law V of *The Basic Laws of Arithmetic*, which states that two functions (concepts) determine the same course of values (extension) if and only if they have the same values for the same argument (the same objects fall under each), taken together with what amounts to having each (concept or "first-level" concept in Frege's sense) determine an extension, which is an object [Frege, 1967, 36-46; see also xii, xxxvii-xl].

Axiom (5) of the (P-D) postulates can be taken in two ways. Looking at (1-4), we see that the implicit (or explicit) quantifiers range only over subject terms, in a first-order schema. We can take (5) as a "postulate schema" with "ϕ" as a predicate (or class) variable — or "schematic" letter — that is not quantified over. Thus, any definite predicate that replaces the variable yields a definite instance of the postulate. We thus have indefinitely many or denumerably many postulates as instances of (5).[5] Alternatively we can read (5) in terms of a predicate quantifier, as (5*): For any ϕ: if 0 is a ϕ & if for any y: if y is a ϕ then the *successor of y* is a ϕ, then every *number* is a ϕ. (5*) is a *second-order* (as that expression is now used) and, in one sense of the term, "impredicative" formula. Where the instances of (5) are obtained by "replacement" of a constant predicate for the free variable, those of (5*) are obtained by universal instantiation. Using (5*) the logicist construal of "*number*" turns the latter into an "*impredicative* predicate" in the sense that it is one of the predicates (represents one of the properties) that the predicate quantifier "ranges over." This will play a role in M. Dummett's arguments against the recognition of an infinite set of natural numbers, which is the basis of his claim that the concept of *natural number* is "indefinite," and that, in turn, is employed in his criticisms of logicism as a philosophy of arithmetic.

1 FREGE

Frege developed his variant of logicism in connection with setting out a philosophy of language and theories of reference, meaning, and predication and his rejection of a kind of "formalism"[6] as well as psychologism in mathematics. Doing so involved him in detailed discussions of sense and reference, concepts and objects, as well as of functions, arguments and "truth." In those discussions he ingeniously used a few basic ideas to attempt to resolve a number of fundamental philosophical problems that, as he rightly saw, paralleled specific problems in the philosophy of mathematics. For example, consider the questions posed by distinguishing between trivial identities, like "7=7," and significant claims, such as "2=the positive square root of 4." It is obviously paralleled by "Socrates=Socrates" and "Augustus=the only adopted son of Caesar." Since "names" like "7" and "Augustus," as well as the descriptive phrases, can be said to be used to refer to or denote a particular individual or object, while what we understand when we use the different expressions is generally taken to be different, Frege took expressions like "2" and "the prime even number," while denoting the same object, to have different *senses* and the sentences "2=2" and "2=the even prime number" to express different *propositions*. In fact Frege took the expressions to denote the same object since the different senses (meanings) that they expressed denoted that object.[7] Denotation was basically a relation between the sense of an expression and a referent, and only derivatively a relation between the expression and the referent. Thus he employed senses of expressions to account for a particular use of a denoting expression being used to denote a certain object and held that "names" like "Pegasus," which did not refer to anything, were meaningful since they had a sense though lacking a referent. Thus, the referent of a sign that did denote an object was not what gave sense or "meaning" to a name. Rather, the expression referred to what it did refer to in virtue of the sense that it was taken to *express*. Having introduced senses, including *propositions* (the latter, being what sentences expressed, Frege called "thoughts") in his account of the meaning of expressions and, hence, also of the differences in meaning, Frege developed a philosophy of language, along with his philosophy of mathematics, to clarify and systematize his ideas.

---

[5] One can have a finitely axiomatizable "version" by using standard order axioms and a relation, such as *is less than*. See, for example, Gill [1990, 1-3].

[6] He mentions the mathematician E. Heine in criticizing those who take mathematics to be a "game," where the numbers become the arithmetical signs and the definitions are construed as rules for moving such objects around, as with chess pieces [Frege, 1967, 10]. He is not here discussing Hilbert, though he would write critically of, and enter into controversy with, Hilbert in later years. Ramsey noted a contrast between the early formalists, mentioning Heine and Thomae, that Frege is criticizing and Hilbert:

> Early formalists all got muddled. They said to avoid all doubt as to nature and existence of numbers let us take them to be signs "1", "2", "3". Then mathematics consists in taking various formulae involving these signs and forming others from them according to certain rules. .... Also said that real numbers were infinite series of signs. *Hilbert* does not commit these absurdities [Ramsey, 1991, 183].

Platonic realism is often associated with the recognition of abstract entities — entities not locatable spatially or temporally. That view, as it relates to mathematics, can be simply presented. Consider a cat lying on a mat. The statement that a cat is on a mat is true in virtue of two objects, the cat and the mat, being in a certain arrangement of spatial proximity — a relation. Thus, for it to be true there must be a cat and a mat. Pictures and stories not withstanding, we do not have a unicorn lying on a dragon's back, as there are neither unicorns nor dragons. Hence, a corresponding statement about a unicorn on a dragon's back will not be true in a straightforward sense. Plato can be seen as holding that given the meaning and the truth of arithmetical statements, like "7 > 2," we must recognize the numbers, 7 and 2, and, implicitly, the *relationship* between them. Extending this simple idea to the case of "Socrates is wise," one can see Plato as moving to the need to recognize *wisdom*, as well as Socrates, in order for it to be true that Socrates is wise, that he *participates* in wisdom — a form or "abstract" object.[8] Such objects, like numbers, are not only abstract as not being spatio-temporal but in another sense. One can think of a common attribute, as one can of a number, apart from (or abstract from) the particular objects it is applied to — as abstracted from them. Thus, while we speak of or see 7 cups, 7 saucers and 7 seven chairs, we can consider the number 7 apart from "concrete" realizations of it — as what the various collections have in common. Hence, we can merely imagine 7 people having tea or 7 unicorns, for that matter. Noting certain similarities between 7 and wisdom, one can see Plato's theory of forms as linked to a concern with mathematical entities, as abstract entities, to account for the truth of mathematical statements. Thus, one can think of Plato, influenced by Pythagorean ideas, seeking correspondent *abstract* entities to "ground" both the meaning of arithmetical terms and arithmetical *truths* — arithmetical objects and "facts" that were not in the spatio-temporal realm.[9]

---

[7] Some understand the senses of signs like 'Socrates' and '7' in terms of the notion of "concept-of" — e.g. concept of Socrates, concept of 7 — as opposed to Frege's notion of a concept. This is reminiscent of supposed "individual concepts."

[8] Once the introduction of the forms led to the explicit recognition of the connections they have with each other and the "ordinary" things that "participate in" or "reflect" them, the stage was set for the long history of the puzzles surrounding "predication" and relations that stretches from Plato's notion of *participation* and Plotinus' process of *emanation*, thru the medieval realism-nominalism controversies, to the empiricist attacks on "abstract" ideas and on to Frege's "unsaturated" functions and Russell's universal relations, qualities and logical forms.

Frege's distinguishing sense from reference led him, at places, to take statements, whose subject terms expressed senses but lacked referents, to lack truth values, while at other places he supplied an arbitrary referent for the term (and thus a truth value for the statement), and at yet other places he suggested dismissing such terms from a scientific language. A further question that arises is whether the sense-reference distinction applies to predicate expressions as well as expressions for objects. Interpreters of Frege have disagreed about this, yet it seems clear that a predicate has both a sense and a denotation, the latter being a concept, though Frege speaks of the concept both as what is expressed and what is denoted by a predicate term. But while it is the sense of a term like "2" that enters into the proposition expressed by "2 is even," it is apparently not the sense of the predicate, "is even," but the denoted concept itself that is the incomplete constituent of the proposition. The sense of the expression referring to the concept plays different roles. One is that the sense seems to be the relevant constituent of a proposition expressed by a sentence like "Socrates falls under *the concept man*," for on Frege's account the function (concept) unifying the proposition is *falls under*, not *is a man*. This gave rise to a problem once discussed by Frege commentators. Frege insisted that an expression like "the concept man" cannot denote a concept but must denote an object, since the expression does not play a predicative role. So, in the case of certain contexts, where the relational predicate is "falls under" for example, Frege suggested simply avoiding such expressions as misleading and due to "...the awkward position in which language here finds itself..." [1967, 37].

Just as Frege took signs like "Socrates" and "7" to have both a sense and a denotation, he took sentences like "Socrates is wise" and "7 is odd" to have a sense, the proposition (or "thought," as he called it) expressed, and a referent. But the referent was specified as the truth value — True, in these cases. Thus the arithmetical and the biographical statement referred to one and the same "object" — the True — as all sentences with truth values denoted either the True or the False. The doctrine sounds strange. It is one thing to recognize "senses" in addition to "referents" as, apparently, further abstract objects, in order to account for meaning. It is another thing to recognize truth values as objects. To be sure, one who reads logic texts is familiar with the occurrences of the signs "F" and "T" and their being said to designate truth values, as one is familiar with the use of numerals to designate numbers. But in logic classes one no more stops to ask if there really exist truth values than one stops in mathematics classes to ask if numbers really exist. Frege explicitly took there to be such objects. This resulted in part from the application of his sense-denotation distinction to sentences as well as to names like "Socrates," in part from his division of terms into incomplete terms representing "functions" and complete terms representing objects and in part from his belief that attempts to define "truth" and purported theories of truth, such as the correspondence theory, were inadequate. Frege wrote:

---

[9] That Plato was both influenced by and concerned with Pythagorean ideas, as in the *Phaedeo*, is often noted.

> Can it not be laid down that truth exists when there is correspondence in a certain respect? But in which? For what would we then have to do to decide whether something were true? We should have to inquire whether it were true that an idea and a Reality, perhaps, corresponded in the laid-down respect. And then we should be confronted by a question of the same kind and the game could begin again. So the attempt to explain truth as correspondence collapses. And every other attempt to define truth collapses too. For in a definition certain characteristics would have to be stated. And in application to any particular case the question would always arise whether it were true that the characteristics were present. So one goes round in a circle. Consequently, it is probable that the content of the word "true" is unique and indefinable [1968, 510].

One can imagine him believing that we might as well take true propositions to denote the truth value True, since it is hopeless to try to provide an account of truth. He may also be claiming that no definition of the predicate "is true" is feasible, since such a definition will have to be of the form "x is true if and only if x is χ ," where χ is some condition or "characteristic" that is had or fulfilled by whatever truth is ascribed to. But, then, one must hold that x is true if and only if it is true that x has or fulfills χ. Thus, we would be involved in another vicious circle. Another claim that he could be making is that to offer such a purportedly explanatory definition (as opposed to a stipulation) is to assert or presuppose that the explanatory statement is itself true. But then we must know what "truth" means in order to offer such an explanation, for we must know that it is true. Thus we cannot explain what it is to be true. Yet another way of construing his argument is to take him to claim that a theory like a correspondence theory proposes a definition of "is true" or claims to "analyze" the notion of truth along the following lines:

1. Given any proposition p, p is true if and only if there is a fact that p corresponds to.[10]

But then, "any" in (1) applies to (1) itself, so (1) is true if and only if there is a fact to which it corresponds, and we have

2. (1) is true if and only if there is a fact that (1) corresponds to.

Hence, by the correspondence theory that offers (1) as an analysis of truth, (2) must be true. But on the correspondence theory, (2) will be true if and only if a further statement holding that it corresponds to a fact is true and so on. Thus, a proposition will be true if and only if there are an infinite number of facts, and we cannot set down all the facts that its being true "presupposes." This argument assumes that the same ground of truth will not suffice as a truth ground for each element of the series of sentences containing "is true" that is thereby generated. That is, it is assumed that each sentence in the series corresponds to, or is "made true" by, a different fact. One can point out that given that 7 > 2, there is no reason to hold that its being so does not suffice as a ground of truth for the series of trivial conjunctions “7 > 2&7 > 2,” “7 > 2&(7 > 2&7 > 2)”, etc., as well as for the series of sentences “‘7 > 2’ is true, ” “““‘7 > 2’ is true” is true,” etc. In both cases we simply have a series of different true sentences.

---

[10] Such a claim thus involves issues related to Russell's "types" and, later, Tarski's semantic levels, and to further questions about the feasibility of a *univocal* truth predicate.

Dummett, in supporting Frege’s attack on the correspondence theory of truth, sees Frege as arguing that, on a correspondence theory, one cannot derive what is known as the Tarskian T-sentence for statements: “‘Socrates is wise’ is true if and only if Socrates is wise,” for example [Dummett, 1973, 444–52, 463–64]. Hence, if one accepts Tarski’s condition for introducing a satisfactory truth predicate into a schema, a correspondence theory supposedly fails. It is of historical interest that both Moore and Russell, impressed by the arguments of F. H. Bradley and B. Bosanquet against the correspondence theory, rejected such an account in the early part of the 20*th* century. Thus Moore took true propositions to be true in virtue of an “inherent property” based on their constituent concepts standing in a basic *true-making* relation, as the constituent concepts of false propositions stood in another, false-making, relation [Moore, 1899, 180-83]. This suggested taking all true (false) *propositions* to be necessarily true (false) as mathematical statements were commonly taken to be. In an 1898 letter to Russell he had written:

> Truth therefore does not depend on any relation between ideas and reality, nor even between concepts and reality, but is an inherent property of the whole formed by certain concepts and their relations; falsehood similarly.

Thus Moore’s taking true propositions to be their own ground of truth is perhaps where the identification of facts with true propositions, by more recent philosophers in the “analytic tradition,” stems from. Russell, in turn, objected to the correspondence theory’s need to appeal to “objective falsehoods” [1906/7, 45-47]. Bradley, interestingly, had his own version of Frege’s argument [1944, 110]. Given Frege’s views on truth, his logicism, and rejection of “psychologism” and “formalism,” in mathematics would not have been motivated by his adhering to a correspondence account of truth.

The connection Frege saw between the use of functional expressions in mathematics and concept expressions in ordinary language led to his considering a concept like *is wise* to coordinate objects, like Socrates, taken as arguments, to truth values, the True or the False. Thus he thought of concepts as functions whose values were truth values, and in fact offered that definition of what a concept is, with predicates being expressions for such functions. Neither the notion of a function nor that of an object (argument) was “defined” by Frege — they remained two of the primitive logical concepts of his logistic analysis — though he characterized functions (and function signs) as “incomplete” and “unsaturated,” while objects (and the signs denoting them) were “complete” and “saturated.” (These characterizations are still reflected in standard terminology as signs like ‘x is wise,’ ‘Fx’ and ‘ϕx’ are classified as “predicate” expressions (as well as being called “open” or “incomplete” sentences), while ‘x’, ‘Socrates’ and ‘2’ are, typically, “subject” signs.) As there are arithmetical functions of two arguments, like $x + y$, so a dyadic relational concept, like *is wiser than*, requires two arguments to yield a truth value as value and is thus logically distinct from monadic concepts.

Frege took a monadic concept, such as *is even*, to determine an extension as its “course of values” — in effect a class of (ordered) pairs consisting of an object, say 2, and a truth value. Thus $\langle 2, T\rangle$ and $\langle 3, F\rangle$ would belong to the course of values determined by the concept *is even* and the extension of the concept is the class of all objects *falling under* the concept — those paired with T. In this way every concept determined another entity, a class, as every concept expression or predicate “χ” gave rise to a class expression “the class of χs.” But classes, unlike concepts, are objects and denoted by “complete” or “saturated” expressions that can occupy a subject place in a sentence. This fit with another fundamental set of ideas Frege had about numbers as objects of a specific kind.

Consider two collections of three objects: three cups and three saucers. *Having three elements* is an attribute of both collections, for both have three elements — as the three cups are all alike in being cups. Take the collections to be classes, whatever it is that classes are understood to be. One can then think of the natural numbers or positive integers as concepts or attributes applicable to classes — of the number three as the concept “has three elements.” That one employs the term “three” in the phrase “has three elements” does not pose the problem of explaining a notion in terms of itself, for one easily construes that use — “there are exactly three things such that . . . ” in solely logical terms along the lines of “$(\exists x)(\exists y)(\exists z)((x \neq y \& y \neq z \& x \neq z \& \ldots x \ldots y \ldots z) \& (\forall w)(\ldots w \ldots \leftrightarrow (w = x \vee w = y \vee w = z)))$” [Carnap, 1937, 143]. The real problem Frege faced was that his basic dichotomies of argument and function in mathematics and, derivatively, of object and concept, require numbers to be objects (arguments) and not concepts (functions). A simple maneuver suggests itself. The concept of being a three-element class determines an extension, a class — the class of all three-element classes. Such a class, like all classes, is an object. This, in turn, suggests taking the number three not to be a concept, such as *having three elements*, but as the corresponding class, the class of all three element classes. This is what Frege later did [1967, 99-100] and Russell followed him in doing. Frege’s earlier version was somewhat different.

In his *Grundlagen*, Frege took the number three not as the class of all three-element classes, but as the class of all concepts that determined three-element classes. And so for all the natural numbers. The 0-element class, the null class, is determined by an empty concept, the concept of not being self-identical. One might find an apparent circularity in taking the natural number 3 to be a class of concepts (or of classes) applying to (containing) three-element classes. But it is not. For one can consider the class of all concepts equinumerous with a given concept — equinumerous in the sense of there being a one-one function holding of them. This does not make use of the notion of a natural number as one is not speaking of any particular number of elements of a class. One simply takes a cardinal number to be a class of all concepts (or classes) equinumerous with some concept (class), as 0 is the class of all concepts equinumerous with the concept $x \neq x$. All one then need do is specify a way of getting the rest of the natural numbers from 0, without appealing to numbers in the process. Having introduced 0, Frege could make use of the concept “being identical with 0.” That concept has one object falling under it — the number 0. He could then consider the class of all concepts equinumerous with the concept being identical with 0. This yields the class of all one-element concepts, which is then taken to be the number 1. Having 0 and 1 we can proceed to 2, via the concept — being identical with 0 or being identical with 1, which applies to two objects, 0 and 1. Of course one cannot proceed step by step in such a manner or simply say “and so on.” But one can characterize the notions of *successor* and of *natural number* by means of "logical concepts" (such as "class" and the standard concepts of logic and set theory) in such a way that the class of natural numbers contains the infinitely many elements in the line of succession starting with 0, and only those elements. The type of procedure is seen at its simplest in a later set-theoretical variant. Take the null class, $\varnothing$, to be the number 0. Then take "the successor of β" to be the class union of β and the unit class of $\beta - \{\beta \cup \{\beta\}\}$. Thus, with 'S0' for 'the successor of 0', 'SS0' for 'the successor of S0', etc., S0= $\{\varnothing\}$, SS0= $\{\varnothing, \{\varnothing\}\}$, etc. The signs '1' and '2', etc. then simply abbreviate 'S0', 'SS0', etc. What Russell did, was take the unit class of the null class, rather than the null class, to be 0 and specifying successor in a more complex manner was then simply use (P-D)-5 to specify what a natural number is. Following Russell's elegant presentation [Russell, 1953], one defines a hereditary class (property) as one such that if x is in the class (has the property) the successor of x is in the class (has the property). An inductive class (property) is then said to be a hereditary class (property) that 0 is an element of (has). And, finally, the property (class) of natural numbers is the property (class) of having all inductive properties (of elements that belong to all inductive classes). Thus, *is a natural number* is an *impredicative* property in the sense that it is an inductive property while specified as the property of having all inductive properties (correspondingly the class an impredicative class).

Frege and Russell thus answer the traditional ontological questions about numbers — Do numbers exist? If so, what are they? The answer to the first is "Yes," for Frege, but "No" for Russell, as we shall discuss later, since Russell did not take classes to exist on his "no class theory" of the first edition of *Principia Mathematica*. Frege's response to the second is that they are classes, hence objects. More specifically one can say: "A (cardinal) number is a class of concepts such that all concepts in the class are equinumerous and any concept equinumerous to a concept in the class is in the class." (What he actually does is closer to (i) defining "the number which belongs to the concept F" as "the extension of the concept 'equinumerous to the concept F"' and then (ii) define "*n* is a (cardinal) number" as "There exists a concept such that *n* is the number which belongs to it" [Frege, 1953, 85].) The second part of his logistic construal of arithmetic was to find a way of showing that arithmetical truths could be logically derived from logical truths, having supposedly shown that arithmetical concepts could be construed in terms of logical concepts. Arithmetic could then be claimed to be a matter of logic.

2 CONTEXT, NUMBER AND NEO-FREGEAN LOGICISM

One of the most widely discussed aspect's of Frege's philosophy of language has been the meaning and significance of his so-called "context principle." For Dummett that principle achieves the status of a philosophical axiom along with the more familiar neo-Wittgensteinian dictum — to know the meaning is to know the use. In fact Dummett takes the two principles to be pretty much the same and to allow him to recognize abstract objects and reject popular causal accounts of reference and meaning. The latter accounts question our being in "causal contact" with abstract objects and, as we shall see, take that to be a ground for rejecting such entities, and hence "logicism," in "nominalist" fashion. Dummett observes that since we know when it is correct to say things like "We crossed the Equator" and know the connection of the Equator to certain truths about the Earth (its having poles, etc.), even though the Equator, unlike the Earth, is an abstract object, we can safely speak of the referent of the expression "the Equator" [1991, 182]. This illustrates the importance the context principle plays in Dummett's own thinking and its link to his focus on the "use" of language. But it also shows that he resolves classical ontological problems, like the problem of universals, by simply noting that there obviously are common properties and referents of words like "red," "white" and "tan" — referents other than the red, white and tan things that "satisfy" the predicates. For we determine the truth values of sentences like "Mondrian used the color red extensively, but never used tan," "Red is darker than white" and "Red is a primary color but tan is not" quite readily. One can take Dummett as holding the problems to be resolved too simply or as not being real problems, since his polemic against "nominalists" is really a dismissal of the question nominalists purport to answer. This point is further illustrated by Dummett's view that Frege was the first to take "the linguistic turn" by turning "an ontological question into a linguistic one." Dummett sees Frege's early attempts to define "cardinal number" as taking that turn for the first time. Hence Frege's logicism was a key to the development of analytic philosophy in the 20^{th} century. To Dummett the significance of Frege's defining "cardinal number" as he did is not taking numbers "to be classes" but specifying a context for numerical expressions that provides a way of verifying the truth or falsity of sentences employing numerical expressions, such as "The number of people in the room is 7."

The context principle may be stated as: Only in the context of a statement does a term (phrase) have a reference and sense. (Those who participate in the discussions about it tend to distinguish taking it with respect to referents and truth values and taking it in terms of senses and propositions.) Whereas Russell would take the meaning of a genuine or logically proper name to be the object it is taken (used) to refer to — the interpretation — Frege supposedly holds that such a term would neither have sense nor reference outside a context supplied by a statement. Ironically, given Frege's later distinction between sense and reference, there is a way in which the context principle can be seen to be both true and trivial. Consider the sentences "2 is even" and "Plato believes that 2 is even." On the view Frege developed after his *Grundlagen* statement of the context principle, the expression "2" does not have either the same sense or the same referent in them, since it "shifts" its sense and reference as it occurs in the second sentence. In this case, the key shift is from the embedded sentence denoting a truth value, as it occurs by itself, to its denoting the proposition that is its sense as it occurs "by itself." Thus in a totally trivial sense the context determines the sense and the denotation. This was Frege's key to resolving problems posed by certain "intensional" (as well as "intentional") contexts. Other complications to Frege's view are posed by the way in which a sentence may be "deconstructed." The sentence "*aRb*" can be "decomposed" in terms of the names "*a*" and "*b*" and the dyadic predicate "*xRy*," or of "*b*" and the monadic predicate "*aRy*," for example.

Some defend the context principle on grounds that connect with the so-called verification criterion or theory of meaning and a once familiar issue raised by various interpretations of Wittgenstein's "private language" argument (purportedly showing that the notion of a private language is incoherent since there are no criteria for determining correct use). Consider, for example, someone uttering a token of the word "Socrates" and suddenly dropping dead before being able to say anything else. We might never be able to figure out what she was going to say. Whether she was referring to the historical figure or her neighbor's dog, named after him. Surely it doesn't follow that she wasn't referring to something on the grounds that she didn't finish the statement or since we cannot determine who or what was referred to. We would simply never know whether she was referring to some person, or other object or simply uttering sounds. (Consider her having uttered "this" or "that" instead of a proper name.) But to even suggest that certainly suggests, by ordinary usage, that she could have referred to something or someone. It is simply obvious that, for whatever purpose one uses the term on such, fortunately rare, occasions, one sometimes has referred (and intended to refer) to someone or something. (Now consider the unfortunate person having been uttering unfamiliar sounds referring to the trees in her garden — the sounds being tokens of the names she has given or is currently giving to them.)

Frege seems to reject the context principle in an 1899 letter to Hilbert:

> It is absolutely essential for the rigor of mathematical investigations that the distinction between definitions and all other statements be respected most thoroughly. The other statements (axioms, basic laws, theorems) cannot contain any word and any symbol whose sense and reference or whose contribution to the expression of the thought did not fully exist beforehand, so that there is no doubt with regard to the sense of the sentence, the thought expressed by the sentence [1975, 62].

In any case, little is to be gained if, like Dummett, one both preserves the context principle for the later Frege (for senses) and advocates it by taking it to state that "... the sense of an expression relates exclusively to its role in sentences, and consists in its contribution to the thought expressed by any sentence in which it occurs. So understood, it is indisputable that Frege continued to maintain it in *Grundgezetze...*" [1991, 184]. That not only fits with what is said in the letter to Hilbert but, depending on which words are stressed, can be taken to turn the principle upside down and into a way of expressing support for a principle of compositionality. Dummett, in fact, entitles a section of his book "A compositional interpretation of the context principle" [1991, 202]. But his discussion is disappointing, as, repeating a theme from an earlier book [1973], he merely claims that a mastery of basic "characteristic sentences" or "recognition statements" is required for our understanding of a term. These are surprising, as well as unhelpful: "... statements of the form 'This is T', where T is the term in question, or 'This S is T' where S is a sortal such as 'person', 'street'..." [1991, 204].

Russell, who sometimes took the meaning of a logically proper name to be given by its referent and that of a primitive predicate to be the property or relation it represented, views Dummett rejects, could easily accept a context principle in such a form, even without accepting propositional entities.[11] Russell's saying, in his logical atomism essays that predicates and names referred in different ways (and didn't exist in the same sense of "exist") reflects that. Nor is any more said by Dummett's explication that "The context principle ... amounts to the conceptual priority of thoughts over their constituents: the constituents can be grasped only as potential constituents of complete thoughts" [1991, 184]. For Dummett seems to be doing no more than reiterating the weak claim that to know the "meaning" (in one sense) of a term is to know its "logical grammar" — what kind of term it is. So that even if one knows the referent of a logically proper name, knows its meaning in the sense of *what* it refers to, one must know *that* it can combine with predicates of a certain kind to form a sentence, that it cannot function as a predicate, etc. One can even stretch this into holding that one knows its "role" in "the language" as "a whole." This is merely a kind of innocuous logical "holism" or contextualism that need not be disputed by those who espouse a reference theory of meaning and seek to avoid the recognition of senses and propositions as entities, while rejecting the replacement of classical philosophical issues by a focus on the "use" or "role" of words and sentences in ordinary language, including ordinary mathematical usage. It is an odd feature of analytic philosophy how such notions, derived from the logical holism of Wittgenstein's *Tractatus*, have been transformed into extensively elaborated and discussed "philosophies of language" within contemporary analytic philosophy. Thus one finds G. Evans announcing a "Generality Constraint" whereby one can be said to comprehend a thought only if one can be said to also comprehend every thought of a certain kind. Such a network of connected thoughts form a kind of limited holistic context. Thus I can be said to understand "6 is even" only if I also understand "7 is odd" [1982, 100 ff.].

---

[11] In a way he did since he often took predicates, in Frege's manner, to be complex expressions indicating a logical structure, as "Fx" and "Rxy" do. Thus predicates and names (the latter by being "simple") indicate their grammatical role [Russell, 1971, 338].

The context principle, as stated in *Grundlagen* [1953, 71-72] is not as clear-cut as some think. For, first, Frege is focused on distinguishing between the "content" of a word and our having an "idea" of it to deny the view that "... any word for which we can find no corresponding mental picture appears to have no content." The way to do this is to avoid taking words in isolation and to focus on propositions, for this will block our thinking in terms of an image or mental picture of some kind. Second, he is concerned with sentences like "The number 0 belongs to the concept F" being taken to indicate that "0" is functioning as a predicate representing a concept rather than being "only an element in the predicate" [1953, 68]. Thus it can still be seen as a "self-subsistent [independent] object." Third, he consistently violates the principle in (a) his discussion of words like "the," as "the definite article claims to refer to a definite object," (b) his noting that the expression "the largest proper fraction" is senseless and "has no content" since no object falls under the concept represented and (c) his saying things like "In this definition the sense of the expression 'extension of a concept' is assumed to be known" (1953, 117). Moreover, in speaking of concepts he constantly seems to presume we understand the meaning of concept expressions, like "is F," and not merely expressions like the one I just used to refer to an expression — the entire expression including the quotation marks — which would supposedly supply the "context" to provide its meaning. Thus, for example, he speaks of a concept being contradictory, though the concept expressions are only mentioned, and not used, in the statements about the concepts. Fourth, he emphasizes the incompleteness of concepts as opposed to the completeness of propositions. This he does by noting that concepts cannot be asserted, only judgment contents can. Thus, in the case of a relational proposition, if you withdraw both terms you are left with a concept that has no more "sense" than a simple concept has. (Austin, interestingly, interjects "assertible" to form "assertible sense" in the translated text [1953, 82].) There is a further point worth noting about Frege's context principle that is related to this focus on concepts as incomplete and his contrasting them with judgment contents. In his 1879 *Begriffsschrift* Frege had already indicated an interesting form of the context principle, one that calls to mind Russell's comments about "incomplete symbols." Frege contrasts concepts like "the number 20" and "every positive integer" as not being of the same rank, as what can be asserted about the number 20 cannot be asserted, in the same sense, about "every positive integer." And he goes on to write:

> The expression "every positive integer" by itself, unlike "the number 20", yields no independent {*selbstandige*} idea; it acquires a sense only in the context of a sentence [1972, 128].[12]

Here it would seem clear that an expression "every positive integer" differs from one like "the number 20" in an obvious and innocent sense of requiring completion to express a definite thought that represents a truth value — hence a definite object, while the latter expression, as it stands, can be understood to express a "*selbstandige*" idea that denotes a specific object, the number 20.

Dummett infers, from his views about meaning and the context principle, that to understand a statement is to have the ability to recognize "what counts as verifying the statement, i. e. as conclusively establishing it as true." This is a generalization of his construal of the "central notion of the theory of meaning for mathematical statements" where the notion of *truth* is replaced by that of *proof* [1978, 225-26]. Understanding a mathematical statement, say S, requires our being able to distinguish proofs of S from non-proofs, and this ability to distinguish proofs from non-proofs cannot then depend on our understanding the meaning of S. For S has no meaning or truth value prior to our distinguishing proofs from non-proofs, and thus appears to be arbitrary. In what sense then is S a statement (or an expression of a statement) as opposed to a mere string of marks? And in what sense can one speak of such a "proof" or "disproof" of S as establishing it to be true or false? Lindström, taking a sentence to be a "string of symbols," argues that one does not prove (or even assert) sentences, in the ordinary sense of "prove," and thus establish "strings of symbols" as true. Thus proofs in Dummett's sense "are not proofs in the ordinary sense; indeed, they have almost nothing in common with proofs in the ordinary sense" [2000, 134]. If we have a statement, we must clearly know what it means to grasp that a sequence of other statements provides a proof of it. Moreover, there is no basis, aside from having chosen them, for adopting one set of rules of logic as opposed to another, especially the rules of classical logic. This has the consequence, on Dummett's view, that the meanings of the statements change if the rules of inference differ, so, for example, the meaning of arithmetical statements in terms of a system of classical logic is different from that given to them in a system employing intuitionist logic. But this poses a problem for Dummett's view, as it would seem to follow that with respect to a system of classical logic we have, as we will later discuss in connection with Carnap, a purely "internal" question regarding whether they "establish their conclusions as true" and hence the answer is trivially "yes." For that follows from the "meaning" of those conclusions. But Dummett takes the answer to be "no," for how else could he reject one system as incorrect and accept another as correct, thus employing "external" criteria, instead of merely proposing the adoption of one linguistic framework rather than another? Since Dummett must then accept that "these conclusions have a meaning independent of what is regarded as (defined to be) a 'proof"', Lindström argues that Dummett's view is therefore not only implausible but incoherent [2000, 135].

---

[12] German speaking scholars suggest that the expression "a sense" should probably be replaced by "its sense."

The "finitism" embodied in Dummett's verificationist approach to logic and mathematcs lead him to reject our purportedly having a definite "idea" of an infinite totality, and hence a definite concept of the natural numbers as an infinite collection. His argument has changed over time. Initially he held that an "indefinitely extensible" concept is ambiguous. Thus, the Russell class of all classes that are not members of themselves cannot be taken to be a definite totality, R. For, since R cannot belong to itself without paradox, we have to recognize that R is not a member of itself and think of an extended class, $R* = R \cup \{R\}$. But then we can "extend" that to a new class and so on. Hence the concept — being a class of classes that are not members of themselves — is "indefinitely extensible" and "inherently vague" as it does not have a determinate extension. What Dummett has in mind is a notion Bernays had employed in a 1935 paper purporting to refute Platonism in mathematics:

> The essential importance of these antinomies is to bring out the impossibility of combining the following two things: the idea of the totality of all mathematical objects and the general concepts of set and function; for the totality itself would form a domain of elements for sets, and arguments and values for functions [1964, 277].

Noting that the concept of natural number was not indefinitely extensible, since the class of natural numbers is not itself a natural number, Dummett sought another ground for holding it to be "indeterminate" or "inherently vague" [1978, 197]. He found it in the fact that the natural numbers form a totality to which mathematical induction applies. For that involves an "impredicative" use of quantifiers and the idea that induction holds for any well-defined property, where the concept of a well-defined property is indefinitely extensible. This is supposedly so since a well-defined property is one expressible with the linguistic apparatus of a precisely specified language. We can then define, "by reference to the expressions of this language" another property "which we also recognize as well-defined, but which we cannot express in the language" [1978, 198].

To speak of properties being language dependent in this way and of "properties," not predicates, as "defined" raises questions, though logicians generally speak of a class being "well-defined" or not. This bears on the early disputes among Russell, Ramsey and Wittgenstein regarding Russell's "Axiom of Reducibility" (see below). That aside, Dummett's failure to find a common feature linking the concept of natural number to "inherently vague" concepts like that of *class of classes that are not members of themselves* leads him to focus on the "impredicative" version of mathematical induction and, implicitly, the Frege-Russell logistic construal of "natural number." For without the latter construal of the concept, the principle of induction is simply a truth about natural numbers and not a part of the explication of the concept. But Dummett obviously felt the inadequacy of questioning the concept of natural number on such a ground, for he later sought to strengthen it by holding that the concept of natural number simply was indefinitely extensible.

The first step was to argue that the concept of *real number* is indefinitely extensible, given the acceptance of Cantor's diagonal argument for there being non-denumerably many real numbers. Dummett sees that argument as having "... precisely the form of a principle of extension for an indefinitely extensible concept: given any denumerable totality of real numbers, we can define, *in terms of that totality*, a real number that does not belong to it" [2002, 27]. Yet, he does not deal with the totality of real numbers but merely with any denumerable totality of real numbers. It is any of those that can be "extended" by the addition of a real number not in the totality. But, since Cantor's argument takes there to be a determinate totality of real numbers "the alternative is to regard the concept *real number* as an indefinitely extensible one" [2002, 27]. This is hardly an argument. Yet, it is on that basis that Dummett rests his claim that the concept of a real number is "indefinitely extensible." And, not stopping there, he notes that the totality of natural numbers is also not a natural number, i. e. is not a finite cardinal, where we "... assume we have a grasp of the totality of natural numbers..." and asks "... but do we?" [2002, 28] The focal question now becomes one of how we decide that we do. For Dummett the answer is simple, it must be capable of being "conveyed to" someone who does not have it: "... a conception of the totality of the natural numbers is supposed to be conveyed to one as yet unaware of any but finite totalities, but all that he is given is a principle ... for passing from any finite totality to a larger one" [2002, 28]. Thus "The fact is that a concept determining an intrinsically infinite totality ... simply *is* an indefinitely extensible one" [2002, 28]. But that is hardly a fact, as he simply stipulates the result he seeks. One can only look for what lies behind the reasoning.

While we have determinate concepts like *three cups*, *four cups*, etc. and *successor of*, the concept of a natural number intrinsically involves the notion of "infinitely many" and hence of an inductive *class*. One thing he could be doing is bringing in the second-order impredicative version of mathematical induction and taking it as part of the explication of the concept of a natural number. Alternatively, he could be simply holding that applying mathematical induction we take "all" to apply to infinitely many and, using induction in mathematical proofs about natural numbers, we thereby take the concept of a natural number to involve the notion of infinitely many. In either case, we must show that the concept of a natural number, involving the notion of an infinite totality or of infinitely many, can be "conveyed" to one who does not have it. Almost echoing Plato's arguments for the doctrine of recollection and Descartes on innate ideas, he holds this cannot be done.

Frege's context principle is also relevant to contemporary discussions of logicism in view of its connection to the emergence of "neo-Fregean" logicism. In his discussion of a definition of "the number of Fs" by means of a purported *contextual definition*, Frege first considers defining "0" *implicitly* as "the number of the concept $x \neq x$. " He does this, in the *Grundlagen*, along with specifying an inductive procedure for arriving at a succession of the numbers *of* concepts. A few pages later he considers an implicit definition in the context of an identity statement: "the number of Fs = the number of Gs."[13] But he then briefly discusses a problem that supposedly arises in the first case, since one has not determined the truth value of statements of identity involving an expression like "the number of Fs" and a name like "Julius Caesar," as in the sentence "The number of Fs = Julius Caesar." In the second case he raises the problem in terms of "England is the same as the direction of the earth's axis" — after having introduced the problem about what the direction of a line is in order to clarify the issue about numbers [1953, 78]. For the discussion of parallel lines and directions can be put "less clumsily" and "can readily be transferred in essentials to apply to the case of numerical identity" [1953, 76]. His point is that while one has specified when the number of Fs is the same number as the number of Gs, or when the direction of line A is the same as the direction of line B, by means of "F is equinumerous to G" and "A is parallel to B," and has thus specified the condition for statements of identity like "the number of Fs = the number of Gs" and "the direction of A = the direction of B" to be true, one has not determined what "sort" of thing a cardinal number is. Hence we cannot determine that England is not a direction of a line or a cardinal number. We supposedly have no way of determining the truth value of the sentences "The direction of the earth's axis=England" or "The number of planets= England." This has become known as the "Julius Caesar problem." The problem, as Frege indicates, "looks nonsensical" since, of course, "no one is going to confuse England with the direction of the Earth's axis; but that is no thanks to our definition..." [1953, 78]. That is, by our ordinary conception of cardinal numbers we know Julius Caesar is not a cardinal number, but on the implicit definition being proposed there is nothing that plays the role of that conception. Just as a formalist does not specify a definite kind of entity as a natural number, aside from being something fulfilling the (P-D) axioms, so such implicit definitions do not rule out identifying a number with Caesar.

The Caesar problem is a symptom of a problem faced by Wright and B. Hale who attempt to employ implicit definitions, not *contextual* definitions, to develop a significant contemporary variant of logicism. Seeking to resurrect and develop Frege's suggested use of implicit definitions, a use he employed *en route* to his development of his logicist analysis, they have set out a philosophy of arithmetic that has been aptly characterized as "neo-Fregean." A significant portion of Dummett's 1991 book on Frege's philosophy of mathematics is devoted to the context principle and the Caesar problem and contains a lengthy critical discussion of Wright's views (as expressed in Wright, 1983) that dispute the claim that a Fregean program of arriving at a satisfactory account of arithmetical entities, truths and concepts can succeed. Wright's view, later elaborated by himself and Hale [2001], makes crucial use of what is known as Hume's Principle: the number of Fs = the number of Gs if and only if F is in one-one correspondence with G. That principle is essentially what Frege discussed, citing Hume, on the way to providing an explicit definition of "cardinal number" [1953, 73-74]. He briefly considered it since it satisfies the familiar mathematician's requirement for introducing objects of a certain kind: that one state a condition specifying when objects of that kind are identical — identity conditions. But the real issue that is raised is not about contextual definitions, for such definitions, as we will see in considering Russell's contextual definitions for class expressions and definite descriptions, may be eliminable explicit definitions. Rather, the issue raised concerns the problematic appeal to so-called "implicit definitions" as providing "analyses." Taking implicit definitions to be a species of "definition," and hence of "analytic truth," Hale and Wright seek to resolve the Caesar problem, in "To Bury Caesar" [2001, 335-96], by focusing on certain purported "facts of the matter" regarding what sorts of objects numbers are. Such facts supposedly show that they are "not trees, tigers, persons or countries," and, moreover, that "Hume's Principle could serve as a partial explanation whose epistemological status ... might rub off on the Dedekind-Peano axioms."

---

[13] On defining identity itself, Frege writes: "Now Leibniz's definition is as follows: 'Things are the same as each other, of which one can be substituted for the other without loss of truth'. This I propose to adopt as my own definition of identity" [1953, 76].

It is a familiar view, and one that arises in many variations, the "structuralism" of Dedekind can be seen as one, to hold that axioms, such as the (P-D) axioms, provide implicit definitions for their non-logical primitive terms, since the latter are not explicitly defined and are the only terms occurring in the axioms that are not logical terms. The problem lies with the notion of an implicit definition. All that is said by its use is that we limit an "interpretation" of the postulates to terms that will render them "true." Thus we can interpret the term "number" in the familiar way in terms of — 0, 1, 2... — or we can take it in terms of any other progression. Or, one can take it to hold of a "system" of otherwise unspecified "abstract" objects — *abstract* in the sense of being obtained by having the axioms suffice to uniquely determine them. (One might take Dedekind to have held this or the view that, given the axioms, there is no need to say anything further — such as that numbers are classes.) Hence, to speak of an "implicit definition" as if it were a definition that gives an analysis of the concept(s) expressed by the "defined" term(s) is misleading, though there is a purely "verbal" link to the notion of a "contextual definition." For those who speak of implicit definitions often think of axioms providing a "context" that gives "meaning" to terms. But Russell's contextual definitions for definite descriptions provide a way of eliminating the contextually defined expression in diverse contexts. Implicit definitions, as Hale and Wright note, do not do that.

Hale and Wright argue that Hume's Principle suffices to "set up number as a sortal concept" [2001, 15], which is held to be all that is required. If we ask why that suffices, we are told that it is because that is what it is to be a sortal concept — to be a concept that is *partially* explicated in terms of having a "criterion of identity" for its application [14, 15]. Thus the "implicit," not "contextual," definition establishes "number" as a sortal concept (actually, if one is careful, what is implicitly defined is "[the] number-of-Φs") on a par with other sortal concepts. As one can then decide that Hale's horse is not identical with Wright's and that Wright's cabbage is not Hale's, one does not have a problem posed by mixing sortal categories and asking if one of the horses is a cabbage. To think that there is a problem about numbers would apparently "rub off" in that one would then have to think that there is a corresponding problem about cabbages and horses. But the familiar problem with being content with implicit definitions has not changed. It remains what Russell noted long ago: "it fails to give an adequate basis for arithmetic" in that it does not enable us to know whether there are any sets of terms verifying Peano's axioms or specify "numbers to be such as can be used for counting common objects." Doing the latter requires the number signs to have “a definite meaning. . . not merely. . . certain formal properties” [1953, 10].[14] We are still not told what numbers are, as Frege does tell us, nor how they are shown not to be entities at all, as Russell purports to do.

Facing the problem, Hale and Wright take Hume’s Principle to provide an implicit definition of an expression for a numerical operator, “the-number-of-Φs,” that enables them to derive the (P-D) postulates in a “suitable system” of second-order logic — using predicate quantifiers. [Again, it is to be emphasized that they do not deal with the term “number” but with the entire expression containing the purported sortal term “number.”] But they do not, and cannot, escape the problems posed by implicit definitions and the Caesar problem. What they do, by the end of their essay, is arrive at the obvious conclusion that people and numbers are different sorts of things, just as Frege had noted that Caesar was no more a number than England was the direction of the Earth’s axis. But, whereas Frege had also noted that that was no thanks to the implicit definition he was considering, Hale and Wright think they can provide an explanation by appealing to Hume’s principle.

Their purported explanation of why human beings cannot be numbers is that the criteria for “personal identity” are not the criteria for the identity of numbers, which is stated in Hume’s Principle [2001, 396]. What that amounts to is that we rule out certain things from being numbers by being of the wrong “category.” If one stops to think about it, this is pretty much what Dedekind’s structuralism amounts to since the identity of natural numbers is given be their place in the ordering imposed by the axioms. Hale and Wright, in effect, simply add postulates to Hume’s Principle specifying the identity conditions for what are taken to be numbers. They, like Dedekind, by so doing, specify the numbers (actually the “concept” *number-of-Φs*) as the things with a certain purported identity criterion. Persons, as one would expect, do not satisfy the condition as that is not how we identify Cicero and Tully as the same person nor Cicero not being Julius Caesar. 2 being the even prime but not being the sum of 1 and 0, Venus and the evening star being the same, etc. are *sorted* differently. Such identities and differences furnish the requisite facts and categorial differences. If we are not willing to accept that the purported problem is resolved by imposing such facts, in the form of establishing categorial restraints that forbid “cross-categorial identifications,” they see no problem. For it is a problem that “afflicts all sortal concepts” [2001, 396]. Hence there is no special problem posed by taking Hume’s Principle to give us the understanding of numerical terms since it “rubs off . . . epistemologically” on the (P-D) axioms — “And that is good enough” [2001, 396]. If you reject Hume’s Principle as being sufficient to provide a philosophical analysis of the problems posed by arithmetical “objects” you are supposedly faced with the insoluble problem of saying why Hale’s cabbage is not Wright’s horse. But that, aside from being difficult to take seriously, overlooks an elementary fact.

People, animals, cabbages and trees are familiar objects with familiar properties. Numbers, as purported objects, pose questions that ordinary physical objects do not pose. (i)

---

[14] At places Russell was less definite and more pragmatic with regards to his own choice of such “definite objects,” for he held that “. . . any indubitable set of objects” that “satisfy the formulae of arithmetic” may be taken as numbers [1952, 209-10]. It is worth stressing his use of “indubitable” — ruling out “abstract” objects determined solely by the axioms — and recalling that by the “no-class” theory of the first edition of *Principia*, number expressions were defined in terms of class expressions, which, in turn, were eliminated by contextual definition and, hence, were not taken to represent anything.

Are there such objects? (ii) If we hold that there are, then, aside from being objects that satisfy the (P-D) axioms, what are they? (iii) Are purported arithmetical truths "really" true and, if so, how do they differ from ordinary factual truths? (Or if they are not really true, why are they not so?") What logicism purports to do is identify numbers as a certain kind of object (Frege) or explain how to avoid taking numbers to be anything at all (Russell) while, for both Frege and Russell, specifying how the truths of arithmetic are to be construed. In the case of trees, tigers, cabbages and people there are neither corresponding questions about whether there are such objects nor about what sorts of objects they are (that there are related scientific questions one can raise is another matter, as are issues about proof procedures in mathematics). Moreover, statements of fact about things like cabbages and kings do not lead to puzzles about their connection to logical truths. That is why logic courses often begin by contrasting true statements of empirical fact with "analytic" or tautological truths of logic. That contrast has a long history. Thus one finds the distinction between necessary and contingent truths made by Augustine, along with his making use of an arithmetical example. Of course one can raise classical philosophical problems about physical objects and our perception and knowledge of them. But one who seriously doubts the existence of physical objects is mad, while one who questions the existence of mathematical "objects" is reasonably taken to have an "inquiring mind." Skepticism about the physical world is simply not the same sort of thing as being puzzled about the existence of sets or numbers — the mysteries of modern physics and the physiology of perception notwithstanding. Thus, typically, one who argues that numbers exist, and are to be taken in a specific way, tends to argue that attempts to avoid acknowledging such objects fail. Supposedly, we are forced to accept numbers, in some fashion, in order to account for certain truths — that there are more people than chairs at a dinner party, laws of physics, that the number of natural numbers is not a natural number, etc. Thus Hilbert's *formalism* emphasized the role of *concrete* sign tokens in mathematics — objects that the mathematician produced, arranged and could observe. And those who deny the existence of numbers often attempt to show how to accommodate mathematical truths without recognizing numbers as entities. Finally, as regards neo-Fregean logicism, there is the general question of the viability of "implicit" definition as a method or tool of philosophical "analysis."

Since the statement of Hume's Principle is taken as an axiom embodying an unexplicated (undefined) use of "number of" that expression is open to various interpretations — and each "unwanted" interpretation must be ruled out — as one rules out "human beings." One has a general way of ruling out unwanted interpretations, in either Russell's or Frege's fashion, by offering explicit definitions of the key arithmetical terms. This is what will ground Frege's dismissal of the "Caesar-England problem" as nonsensical. But even with purported explicit definitions, there are further problems that all forms of logicism face. Russell and Frege each provided a specific interpretation of the three primitive terms in the Peano postulates. Such an interpretation, providing a model, is simply one of many possible interpretations of the postulates [Hochberg, 1956; Benacerraf, 1965]. One who offers such an "explicit" interpretation thus faces the problem of arguing why a specific interpretation is suitable to resolve the philosophical problems posed by arithmetic — What are numbers? What kinds of truths are arithmetical truths? Hale and Wright do not get to the point of answering the first question since they do not offer a specific interpretation in that (a) "number" remains a primitive term as part of a primitive function phrase — (*the*)*-number-of-Φs* — what they call the "numerical operator," and (b) they add an axiom, Hume's Principle, containing that phrase to a system of logic. The axiom is not, as it stands, a statement of logic, but they suggest that it is since it is "...analytic by virtue of being determinative of the concept it thereby seeks to explain" (2001, 14). Thus Hume's Principle, stating an implicit definition, is analytic, and hence a matter of logic, since an implicit definition, being "determinative of the concept" and thus explanatory (i. e. a form of definition), is declared to be analytic. (What is being determined here is the use of the term "analytic.") Such "definitions" are said to be "determinative of" concepts since they provide a way to "fix the meaning of an expression by imposing some form of constraint" [2001, 117]. Since this simply amounts to taking it to be analytic since it is added as an axiom, one can take Dedekind to have done the same by taking a set of axioms, rather than one, to be determinative of a *system* without further specifying just what, if anything, the elements are. Overlooking the fact that just what the entities are, aside from being declared to be *the unique set of entities fulfilling a, or a set of, axioms*, has still not been specified, one might suggest that if specifying "constraints," rather than identifying the entities, eliminates all interpretations but one, they provide a kind of explicit definition by *abstraction* in Dedekind's manner. Wright and Hale believe they have a way of responding to this by holding that "number" is the only sortal concept for which a one-one correlation supplies the criterion for determining identity. In dealing with numbers we deal with a sortal concept that has a unique method for determining the truth of statements of identity — establishing a one-one correlation between appropriate terms. Caesar and England are not "objects" where statements of identity concerning them are explicated in terms of establishing such a correlation.

Again, the similarity to what Dedekind can be taken to have done is obvious, since his numbers are uniquely determined by their place in the progression. Hence the number of Fs is the number of Gs if and only if the one occupies the same place as the "other." What is more, the number of Fs will be found to be **n** if one "counts" the Fs and arrives at **n**. In fact in purely arithmetical identities, like "4 = 2 +2," one does not identify the "number-*of*- Fs" with "the number-*of*- Gs" but, rather, "the number x" with "the number y." The *criterion* for such an identifications is not at all furnished by a one-one correlation but by the place of x and y in the "progression" of the numbers — the "ordinal place" if one chooses to put it that way.

Frege's point regarding the Caesar problem is missed by the "neo-Fregeans." Frege is not talking about identifying the natural numbers with people, he is simply talking about one number being identified with Caesar to emphasize that we do not know what that number is, if we merely characterize it in terms of "the number of Fs," for some concept F. It could still be Caesar, or just about anything. Then Caesar would not be the number of Gs, if G was not equinumerous with F, but would be if it was. All Hale and Wright do is furnish an equivalence, a description of a characteristic of numbers that, in a circuitous way, does what Frege did. For Frege used that characteristic to "identify" numbers as classes of equinumerous concepts. What Hale and Wright do is retain the phrase "number-of-Φs" as a primitive sign pattern. In so doing they do not employ a sortal concept, *number*, in Hume's Principle, just as they do not talk about identity with respect to "the number x" but, rather, with respect to "the number of Fs." Doing the latter, they make use of a function expression, "the-number-of-Φs," in order to introduce the term "number" via such a functional expression. Aside from any problems that arise from the notion of a sortal concept, the specific problem they face arises prior to their having such a purported concept. Thinking about what they do, one realizes, as Frege did, that all one does with Hume's Principle is introduce a condition for when two concepts, say F and G, are said to be equinumerous. One simply then says that the same expression of the form "the number of Φs" applies to both concepts. All one has actually said, then, is "F is equinumerous to (has the same number of elements as) G if and only if F and G are in one-one correspondence" and "the number of Fs equals the number of Gs" merely abbreviates "F is equinumerous to G."

What Frege realized, as becomes clear from the development of the *Grundlagen*, is that Hume's Principle sets out an explicit definition of the relational predicate "is equinumerous to," which is a "second level" predicate that takes predicates, like "F" and "G," as apparent terms ("apparent" indicates that "F" and "G" really occur as predicates when "F is equinumerous with G" is "expanded," just as they do with regard to the Fregean second level predicate "$(\forall x)(\phi x \rightarrow \psi x)$," which, for arguments "F" and "G," yields "$(\forall x)(Fx \rightarrow Gx)$"). Having defined such a relational predicate, Frege was in a position, as Russell would later be, to explicitly define signs that could be used to interpret the three primitive non-logical signs of the (P-D) postulates.

Aside from questions about the viability of implicit definitions as analyses, there is a real philosophical problem with not really using the term "number" in Hume's Principle but, instead, employing the form of expression "the-number-of-Φs" (with "Φ" as an appropriate predicate variable). This introduces a function expression that combines with predicate expressions and is, implicitly, taken to represent a function that takes concepts as arguments. For example, take the concept word "F." Hale and Wright form the expression "the-number-of-F's" from "F" and the "numerical operator" expression that represents a function. But now they face a fundamental problem about "identity." Just ask: What is the function that is represented? Suppose one replies that it is the function that, for every concept taken as argument, maps that argument onto a (cardinal) number as value. But what is a number — i. e. what is the "concept" that gives the elements of the class of values for the "range" of the function? The only answer they have is: A number is what is a value of the function! (Recall that a number is "implicitly" what is a number-of-some-concept.) Thus to specify (identify) the function we must have determined what a number is. But Wright and Hale supposedly specify the "sortal concept" *being a number* in terms of what function the-number-of-Φs is.

There is a further problem. We supposedly have a function coordinating concepts to numbers. Yet a number is implicitly construed (constructed) from a concept and an "operator" that combines with the concept — a number forming operator, one might say. Thus the crucial function, the number-of-Φs, is construed in a two-fold manner — as a function that coordinates a number to a concept and as an operator that forms a number from a concept.[15] The former, key use, clearly presupposes a domain of numbers to be values of the function, while the latter is completely overlooked.

---

[15] The two functions are blurred together as Frege blurred corresponding functions in his discussion of "nega-

What Hale and Wright do is aided by the fact that the standard constructions of elementary arithmetic by logicists, post *Principia,* have been in terms of set theory, where one typically makes use of two primitive notions — "set" and "membership" — and specifies axioms involving those notions. Since "set" and "ε" are generally taken as primitive, one might think that there is little point in offering explicit definitions of the primitive terms of the Peano postulates in terms of the concepts of set theory. There is a "yes" and a "no" to this. However one treats the primitive concepts of set theory, one has still, by interpreting the primitive terms in the Peano postulates and deriving the interpreted postulates as theorems, offered an explicit interpretation of "arithmetic" relative to set theory. On the other hand, there are philosophical problems raised by the concepts of "set," "class" and "membership." Are there sets? And, if so, what are they? Does "ε" represent a relation? What sorts of truths are the truths of set theory? But the fact that such questions remain does not mean we cannot distinguish between offering explicit definitions, as Frege and Russell and other logicists do, and what Hale and Wright do. Frege's dismissal of the "Caesar-England problem" can thus be seen as ultimately based on Caesar and England not being classes (sets) however one construes classes.

3 RUSSELL'S LOGICISM AND THE "NO-CLASS" THEORY

In the introduction to the second edition, 1937, of *The Principles of Mathematics*, Russell noted that, while the book "was published in 1903, and most of it was written in 1903" [1956, 5], the "fundamental thesis" of the book — that mathematics and logic are "identical" — was one that he had "never seen any reason to modify." And he proceeded to discuss two fundamental objections that had arisen in the early decades of the 20th century. One was that of the *formalists*, who, as "led by Hilbert," (see [Hilbert, 1964]) leave "the integers undefined" but assert:

> ... concerning them such axioms as shall make possible the deduction of the usual arithmetical propositions. That is to say, we do not assign any meaning to our symbols 0, 1, 2 ... except that they are to have certain properties enumerated in the axioms" [1956, 5-6].

Thus, as Russell saw it, for formalists, the numerals "do not represent one definite series, but any progression whatever" — what is typically referred to as an ω-sequence. Russell leveled his standard, brief, criticism, that formalists "forget" that numbers are used for counting and that, since "0" can be taken as "any finite integer without thereby making any of Hilbert's axioms false," we lose the familiar *interpretation* of "There were 12 apostles."[16]

---

tion" as a function [Frege, 1970].

[16] What Russell means is that starting with any finite integer as the initial or 0 element one has an omega sequence. One can see what Russell opposes in terms of the following.

> The leading idea of Hilbert's theory of proof is that ... classical mathematics involves an internal closed procedure which operates according to fixed rules... and which consists basically in constructing successively certain combinations of symbols which are considered "correct" or

> The formalists are like a watchmaker who is so absorbed in making his watches look pretty that he has forgotten their purpose of telling the time, and has therefore omitted to insert any works [1956, 6].

In *Principia* class abstracts parallel the contextual definitions for definite descriptions with

$$20.01 \quad f\{\hat{u}(\psi u)\}. =:(\exists\phi)\colon \phi!\mathbf{x}. \leftrightarrow_{\mathbf{x}} .\psi x : f\{\phi!\hat{u}\} \qquad \text{Df.}$$

and

> The condition corresponding to
>
> $$E!(\iota x)(\psi x) \text{ is } (\exists\phi){:}\phi!x. \leftrightarrow_{\mathbf{x}} .\psi x,$$
>
> which is always satisfied because of 12.1" [1950, 188].

The condition, transcribing "∃!" via "exists," can virtually be seen as emphasizing the non-existence of classes, since the apparent statement that a class exists simply states that there is a function equivalent to the "defining" function. [12. 1 is the axiom of reducibility for functions of one argument.] What Russell says about "the existence of classes" is of further interest. The opening sentence introducing the discussion of classes in Section C of *Principia* is: "The following theory of classes, although it provides a notation to represent them, avoids the assumption that there are such things as classes." (*Principia*, 187) One avoids such an assumption due to the contextual definitions for class expressions "just as, in *14, we defined propositions containing descriptions." This is the idea of the "no-class" theory. It leads to:

> When a class α is not null, so that it has one or more members, it is said to exist. [This sense of "existence" must not be confused with that defined in *14: 02 .] We write "$\exists!\alpha$" for "α exists." The definition is
>
> $$*24:03 \qquad \exists!\alpha. = .(\exists x).(x\varepsilon\alpha) \qquad \text{Df. } (\textit{Principia}, 216)$$

The reference to *14: 02 is to the definition of "*E*!" for definite descriptions. This idea had already been expressed in *The Principles*. "Another very important notion is what is called the *existence* of a class — a word which must not be supposed to mean what existence means in philosophy. A class is said to exist when it has at least one term" [1956, 21].

As noted earlier, classes, and class expressions, like definite descriptions, and what they purportedly describe, are both characterized as "incomplete symbols." In the case

---

"proved." This construction procedure, moreover, is "finitary" and directly constructive [von Neumann, 1964, 50].

2 +3 = 3 + 2 is intended to communicate the fact that 2 + 3 and 3 + 2, when abbreviations are taken into account, are the self-same numerical symbol, viz., the numerical symbol 11111. Similarly 3 > 2 serves to communicate the fact that the symbol 3, i. e. 111, is longer than the symbol 2, i. e. 11; or, in other words, that the latter symbol is a proper part of the former [Hilbert, 1964, 143].

of the signs this means they are eliminable by contextual definition and do not have a meaning on their own. In the case of a "purported" represented object, it would mean that the supposed entities are not taken as existents — entities that are assumed to exist by the employment of the linguistic system. Given Russell's no-class theory — his view that classes are not recognized as entities, the *Principia* analogs of the primitive terms in the (P-D) axioms, being explicitly defined in terms of class signs and the remainder of the logical apparatus, are not taken to represent entities — either classes or any other purported "objects." Russell provided a summary statement of these themes in a 1924 essay:

> This definition ... avoids the inference to a set of entities called 'cardinal numbers', which were never needed except for the purpose of making arithmetic intelligible, and are now no longer needed for that purpose.
>
> Perhaps even more important is the fact that classes themselves can be dispensed with by similar methods. ... all the propositions in which classes appear to be mentioned can be interpreted without supposing that there are classes [1971, 327].

4 CARNAP'S *INTERNAL* LOGICISM

The attempt to "construct" arithmetic in *Principia*, and Whitehead's notion of "extensive abstraction" as a tool of analysis, led Russell to reject his Kantian hypothetical or "structural" realism, as set out in *The Problems of Philosophy* of 1911.[17] That realism was replaced by an attempt to sketch a construction of the objects and the spatio-temporal framework of the everyday world out of the directly apprehended data of experience in Russell's 1914 *Our Knowledge of the External World* and subsequent *Analysis of Mind* of 1921 and *The Analysis of Matter* of 1927. As numbers, like classes, were "logical constructions" in *Principia Mathematica*, so physical objects and perceiving subjects would become logically constructed "objects," but constructed from an empirical basis by logical techniques, including Whitehead's "extensive abstraction," and not on a basis of logical concepts alone. Russell's shift from his hypothetical realism about the physical world in 1911 might be considered as an early form of "logical empiricism." For it was this "constructionist" approach to the empirical world of the 1914 book, combined with the logicism of Frege and of Russell and Whitehead, that was a major influence on Carnap's first major book, *The Logical Structure of the World.* That book interjected the term "*Aufbau*" into the philosophical vocabulary and, for a time, epitomized the movement known as "logical empiricism" or "logical positivism" and the group labeled "the Vienna Circle." Carnap, who for some years was an intellectual focal point of "the Circle," was clear about his debt in the preface to the second edition of his book:

> Through the influence of Gottlob Frege, under whom I studied in Jena, but who was not recognized as an outstanding logician until after his death, and

---

[17] This is quite akin to structuralism, as here attributed to Dedekind. For Russell took the realm of physical objects to consist of "unknown" hypothetical objects, properties and relations that shared a "structure" with the realm of directly experienced phenomenal objects, properties and relations (see [Hochberg, 1994]).

> through the study of Bertrand Russell's work, I had realized, on the one hand, the fundamental importance of mathematics for the formation of a system of knowledge and, on the other hand, its purely formal, logical character to which it owes its independence from the contingencies of the real world. These insights formed the basis of my book. Later on through conversations in Schlick's circle in Vienna and through the influence of Wittgenstein's ideas they developed into the mode of thought which characterized the "Vienna Circle." This orientation is sometimes called "logical empiricism" (or "logical positivism") in order to indicate the two components.
>
> In this book I was concerned with the indicated thesis, namely that it is in principle possible to reduce all concepts to the immediately given.... I could carry out my task thanks only to the modern logic which had been developed in the preceding decades, especially by Frege, Whitehead, and Russell... [1967, vi-vii].

Carnap's logical positivism led to the dismissal of traditional philosophical problems and to the rejection of the diverse forms of logicism apparently separating Frege and Russell. For questions about the existence of classes and of numbers were dismissed as pseudo-problems. As he put it in the 1937 English edition of *The Logical Syntax of Language*, the statement that "Numbers are classes of classes of things" was a statement in the "material mode of speech" that led to a "*disregard of the relativity to language of philosophical sentences...*" [1937, 299]. Such sentences really represented "suggestions, " not assertions. The corresponding "syntactical sentence" in the "formal mode of speech" would be "Numerical expressions are class-expressions of the second level" [1937, 300]. Employing transcriptions into the formal mode of speech not only avoided fruitless controversy but allowed the original pseudo-theses to become eligible for discussion and acceptable or not on the basis of their fruitfulness. It did, however, require the specification of a linguistic system for such "theses" to be completed. The latter were thus shown to be relative to a linguistic framework. This emphasized that there was no point (even sense) in disputing whether numbers were objects, whether classes exist or whether numbers are really classes.

Later, focusing on problems of meaning, denotation, truth and modality, Carnap stressed *semantics* in addition to *syntax*. This led to his early work in modal logic and a trilogy: *Introduction to Semantics* [1942], *Formalization of Logic* [1943] and *Meaning and Necessity* [1947] along with the well known paper "Empiricism, Semantics and Ontology." In these works he developed what we may characterize as "internal logicism." This went in tandem with his emphasis on the distinction between "external" and "internal" questions as a basis for the dismissal of philosophical pseudo-questions about mathematical objects — since such questions were meaningless "external" questions. The shift of emphasis took such questions to be replaced by syntactical *and* semantical questions *about* linguistic schemata and the focus was on the pragmatic reasons for choosing one such linguistic schema rather than another.

Contrasted with meaningless external questions that lacked cognitive content, internal questions, like "Is there an even prime between 3 and 8?", could not only be answered within the context of the appropriate system but could be answered simply and "trivially," as the answers would follow from the "rules." Thus the answers to such questions are "analytic," or "logically true" — L-True [Carnap, 1956, 217-218]. The problem Carnap faced in dismissing "metaphysical" questions in general, and philosophical questions about logic itself, was the general problem faced by the positivists in connection with their statement of a verification criterion of meaning. Just as the verification criterion is problematic as applying to itself, Carnap has to appeal to, or take for granted, familiar logical concepts and rules and some system of logic.

Carnap takes the answers to the internal questions about arithmetic to be answered by logical analysis based on the rules governing the expressions. As such rules appear to amount to a set of axioms and rules of inference that comprise a formal theory, one can argue that Carnap's view must presuppose that the formal theory could be a complete formal theory for the system of natural numbers in order to be viable. But, by Gödel's celebrated theorem, it cannot be such a theory. (Thus the problem that many raised for Hilbert's "program" has been raised for Carnap's construal of arithmetic.) Moreover, there is a second problem. As Russell and Frege argued in their attacks on formalism, a philosophy of arithmetic (of classical first order number theory) must recognize that we have to assign the right truth value to each arithmetical statement. But, on Carnap's approach, there is no reason to accept the answers yielded by the system of rules as assigning the right truth values — as being "correct."

> What are these "rules"? The answer seems to be: axioms and rules of inference, in other words, a formal theory. But, the concepts required to define a linguistic framework (formal system) suffice to define an ω-sequence and so Carnap's ideas, at least in the case of first order arithmetic, represent no conceptual gain, or reduction, at all. Furthermore, by Gödel's theorem, there is no formal theory that answers correctly all (first order) internal questions about the natural numbers. Finally, it is not clear why we should accept the answers given by the "rules for the new expressions" as correct; these answers may even be inconsistent. It follows that Carnap's idea (taken as an account of mathematics as it is ordinarily understood and applied) is untenable [Lindström, 2000, 135-136].

As Lindström notes, his line of argument is found earlier, and in greater detail, in Gödel's unpublished "Is Mathematics Syntax of Language?" (eventually published in Gödel, 1986/95) that was originally intended as his contribution to the Carnap volume in *The Library of Living Philosophers*.

In that paper Gödel set out two arguments relevant to the present discussion. The first is:

> Moreover a rule about the truth of sentences can be called *syntactical* only if it is clear from its formulation, or if it somehow can be known beforehand, that it does not imply the truth or falsehood of any "factual" sentence (i. e., one whose truth, owing to the semantical rules of the language, depends on extralinguistic facts.) This requirement not only follows from the concept of a convention about the use of symbols, but also from the fact that it is

> the lack of content of mathematics upon which its a priori admissibility in spite of strict empiricism is to be based. The requirement under discussion implies that the rules of syntax must be demonstrably consistent, since from an inconsistency *every* proposition follows, all factual propositions included [Gödel, 1986/1995, 339].

The point behind the argument seems to be that by Gödel's Second Incompleteness Theorem one will have to justify the rules by making use of mathematics that is beyond the scope of the rules in question — by characterizing the syntax of the system, involving notions like *well-formed formula of the system, proof in the system*, and proving the consistency of the system in a mathematically "richer" system. Gödel seems to be claiming that on Carnap's view one is required to prove the consistency of the rules of syntax, since (i) it is claimed that a rule about the truth of the sentences will be syntactical, (ii) in order to be purely syntactical no true sentence can be "factual," but (iii) if they are not so, "factual propositions" will follow. Thus Carnap's claim that "mathematics" can be taken to be embodied in such rules cannot be true. A less far reaching claim he is making might simply be that if the formal system, embodying the rules of the language, cannot be characterized in a certain way, then it cannot be sensibly claimed that factual propositions do not follow from such rules as the rules cannot be claimed to be formal (syntactical) rules without employing a suitable amount of mathematics in the characterization. Since the rules of the system in question cannot be shown to be consistent, if they are, it would then have to be clear from the very formulation of the rules that they embody no factual content.

The second argument depends on the First Incompleteness Theorem, since, by it, Carnap's notion of something being a consequence of the axioms of the system cannot be taken in terms of it being derivable within the system from the axioms and rules, as not all relevant truths are derivable in any such deductive system. One is forced to take the notion of a "consequence" in terms of a meta-linguistic system of concepts that makes use of a suitably rich mathematical framework.

> ... in order not to substitute for intuitive mathematics a "syntax" as unacceptable for empiricists as intuitive mathematics itself ...it will have to be required that: (1) not only in the rules of syntax, but also in the derivation of the mathematical axioms from them and in the proof of their consistency only syntactical concepts ... be used ... and (2) only procedures of proof which cannot be rejected by anyone who knows how to apply these concepts at all.
>
> The necessity of the first requirement should be beyond dispute. For, if mathematical intuition and the assumption of mathematical objects or facts is to be dispensed with by means of syntax, it certainly will be ... based on considerations about finite combinations of symbols. If, instead, in the formulation of the syntactical rules some of the very same abstract or transfinite concepts are being used — or in the consistency proof, some of the axioms usually assumed about them — then the whole program completely changes its meaning and is turned into its downright opposite: instead of clarifying

> the meanings of the non-finitary mathematical terms by explaining them in terms of syntactical rules, non-finitary terms are [used] in order to formulate the syntactical rules; and, instead of justifying the mathematical axioms by reducing them to syntactical rules, these axioms (or at least some of them) are necessary in order to justify the syntactical rules (as consistent) [Gödel, 1986/1995, 341-342].

Aside from Carnap's views regarding issues posed by first order arithmetic, given his use of the distinction between internal and external questions to reject philosophical problems and, especially, philosophical issues raised by logic itself [1956, 213-217], his philosophy of arithmetic, like Dummett's, raises the familiar question as to whether the logical positivists arbitrarily dismiss, rather than resolve, the issues by labeling them *pseudo-issues*. For Lindström's and Gödel's attacks on Carnap raise the issue of whether what is an *external* question, either about the nature of the "objects" or about what kinds of truths the arithmetical truths are, and hence not a question answerable within a formal system of arithmetic, is an *internal* question within a framework where we ask and seek to resolve *philosophical* questions about arithmetic. In such a context we do not employ procedures of the mathematician or logician, for the questions are not, in Carnap's terms, "scientific." But all this means is that they are not resolved within the context of a formal system that derives such answers by formal rules from "axioms." Such a context would hardly help resolve philosophical issues, however it might help clarify presuppositions various philosophical positions depend on. For there is no canonical method of deciding philosophical issues in the sense that there are such methods for deriving standard arithmetical theorems, and even theorems like Gödel's incompleteness results.

5 CONTEMPORARY CRITICISMS OF LOGICISM

Dummett has claimed that Frege's logicism is inherently flawed since it involves recognizing a domain of abstract mathematical objects while supposedly deriving the truths about the natural and real numbers from logical principles alone [2002, 21]. All it amounts to is the claim that there is an incompatibility between construing numbers as abstract objects in logistic terms (as classes, functions, etc.) and taking arithmetical truths to be logical truths. For logical truths supposedly make no ontological claims — such as, that classes exist. But that depends on how we construe logic. Do the logical truths "$(\exists x)(\exists\phi)\phi x$" and "$(\exists x)(\forall\phi)[\phi x \vee \neg\phi x]$" involve "ontological commitments"? Do the interpretations of the logical signs (the quantifier signs and negation sign in particular) and patterns involve such commitments? Does the appeal to certain domains in such interpretations involve commitments to classes, the elements of the domain, etc.? The answers are neither simple nor easy.

As the term "logicism" has been generally used to apply to Frege's view in the *Grundlagen* as well as to Russell's early form of logicism in the first edition of *Principia Mathematica*, both of those views are often taken to involve an appeal to a domain of abstract objects that numbers are identified with — classes (or extensions of concepts) of a certain kind. Thus it is worth reiterating that Russell's logicism of the first edition of *Principia*, employing his “no class” theory, clearly involved avoiding “ontological commitments” to such objects. Russell eliminated class abstract expressions by contextual definitions along the lines of the elimination of definite descriptive expressions (which we shall soon consider). Thus no “ontological commitments” to classes were involved. Dummett simply seems to assume that “logic” does not involve ontological issues or “commitments” (aside from his ignoring Russell’s variant of logicism, that rejects both classes and numbers as entities). Hence, logical truths cannot, explicitly or implicitly, involve reference to specific entities. One cannot object on the grounds that “Socrates is wise or Socrates is not wise” is a logical truth that contains a term “referring to” Socrates. For it is a logical truth in that it is an instance of “For any $x, \phi : x$ is ϕ or x is not ϕ.” It expresses what Kant would have classified as an “impure analytic a priori” judgment — containing a term expressing “empirical” content — as contrasted with the “pure” universal generalization. (Such “impure” judgments will later lie behind the claims that there are “analytic *a posteriori*” judgments.)

While Dummett objects to the supposed reference to “entities” — to logicism purportedly involving ontological commitments — a less sweeping indictment of logicism focuses on the supposed “kinds” of entities introduced — classes as “abstract” entities not locatable in the physical spatio-temporal realm. This has led, in recent years, to the so-called “problem of access” posed by abstract entities that Lindström has taken seriously. The purported problem stems from the reliance on causal theories of reference and from professed puzzlement about our ability to “abstract” and apprehend “general ideas.” As many have observed, our ability to comprehend abstract ideas led Plato to his version of the doctrine of innate ideas — the Platonic doctrine of “recollection.” It may also have inspired Augustine’s focus on divine illumination, led various medieval philosophers to rely on divine ideas and culminated in Descartes’ explicit doctrine of innate ideas that, on a standard historical perspective, gave rise to the empiricist attacks and the development of an implicit empiricist “meaning criterion. From Locke, Berkeley and Hume to the Vienna Circle, to the more recent contextualist-holism of W. V. Quine, W. Sellars and D. Davidson, and on to much contemporary discussion, one finds the mixtures, in various combinations, of causal theories, of nominalism and of materialism, mixed with attempts to “be scientific.” This is illustrated in Benacerraf’s attacks on logicism and realism about abstract entities.

Benacerraf finds Gödel’s use of the terms “perception”’ and “intuition” unclear when Gödel writes:

> ...we do have a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don’t see why we should have less confidence in this kind of perception, i. e. in mathematical intuition, than in sense perception, which induces us to build up physical theories... [1964, 271].

Benacerraf comments: “...what is missing is precisely what my second principle demands: an account of the link between our cognitive faculties and the objects known” [1996, 26]. His objection is echoed by D. Lewis’s attack on properties, such as colors:

> ...I have complained about the difficulty of understanding the relation that

> allegedly relates concrete things to the abstract simple possibilities — propositions or properties — that they realize... Its a nasty predicament to claim that you somehow understand a primitive notion, although you have no idea how you could possibly understand it [1991, 35-36].

If we ignore the haziness of the notion of an "abstract possibility" that Lewis interjects, what is going on is precisely the line of thought that led British empiricists to reject our having "general ideas" and to their primitive "imagist" conception of thought. While there is a difference between the suspect apprehension of an "abstract" quality, say a color shade, taken as a perceived phenomenal quality that one perceives in perceiving an object of that color, and the apprehension of sets, or other mathematical objects, both are rejected by nominalists of the Lewis-Benacerraf kind along the lines that led Hume and Berkeley to reject abstract ideas.

Suppose, however, we make an effort to consider what we are told we can't consider, two colors (shades of color), say black (B) and white (W) and the truths that one is darker than the other and that Lewis prefers B to W. Assume we take B and W to be universals. They are, like black and white colored patches, simply objects presented in experience. There is nothing mysterious about that, so long as one does not confuse the claim that universals are presented or directly apprehended with the claim that we are not merely presented with the abstract objects or universals but with *the fact that they are universals*. That is another matter. One can argue about whether universals are spatial objects. But if one holds that they are localizable then one gives up certain standard laws that apply to particulars, such as not being at different places at the same time, and certain truths about spatial relations — that *left-of*, as occurring in the perceptual field, is neither reflexive nor symmetrical. It is simpler to take universals as neither spatial nor temporal, while recognizing that they are instantiated by spatio-temporal objects. So what it comes down to is really not a question about mathematical realism or mathematical truth, but the classical problem(s) posed by *universals* and the complications introduced by an extreme form of nominalism being combined with a causal theory of reference and meaning.

There is a simple point to be made about commitments to a causal account. One must wonder what one who advocates such a view takes causality to be. Benacerraf, for example, casually speaks about a causal relation connecting our knowledge or cognitive activity and the (intentional) objects of such knowledge and activity. But how does he construe a causal relation? Though he speaks of a causal relation, he explains neither what he understands by "a relation" nor how a relation (or any relation, as opposed to the terms standing in it) is in physical space-time. This is typical of physicalists. Suppose, however, that there is not a causal relation that in fact obtains between things or events or universals, etc. That is, suppose one adopts a Humean style regularity account of causality. Then the question becomes: How can the existence of an abstract entity be regularly connected to the apprehension of such an entity? The answer is then obvious: the existence of such an entity is a "necessary" condition for its apprehension — where imagining is not a form of apprehending. There is no difference between the black truffle Benacerraf speaks of and the universal black or the set of natural numbers, if such there be. In the case of the black truffle there is, in ordinary terms, the context provided by the physics and physiology of perception in our seeing such an object. But even there, we should distinguish, as

Benacerraf apparently does not, the truffle from the fact that it is black. (One might recall here that Davidson dismisses facts since they cannot be located in space. In the early 1900s Russell and Moore took facts not being spatio-temporal "objects" to be a reason not to speak of them as "existing." Thus they simply, like universals, *had* "being.") And it is not clear how the distinction between the object, the property and the state of affairs consisting of the object having the property, crucial to numerous philosophical attempts to resolve various issues, finds its proper place in the ordinary causal account that explains what happens when we see a black truffle. How does the property black, which, let us assume, is not a spatio-temporal object, causally interact with the eye? How does the state of affairs, which is also not sensibly located in space, do so? And, of course, the causal account stops with physical connections, such as light rays reflecting or being absorbed by an object. Phenomena are simply omitted, yet are relevant. Consider the case of an after image.

An after image is not localized in physical space, unless one identifies it with an occurrence in or a state of one's brain. Yet it may be said to be a partial cause of a memory of having had such an image, if one, say, reports the annoying persistence of such things to a doctor. But how does the after image enter into such a causal chain as it is not in physical space? Perhaps it does so via its assumed correlation with states of the central nervous system that are so localized. But, then, obviously, something like that also takes place when one apprehends a universal quality or thinks of a natural number or grasps an arithmetical or logical truth. If such connections are the key to causal interaction, whatever one means by causality, then there is no mystery about our causal interaction with arithmetical entities and truths over and above the questions about how we come to have general ideas and the perplexities of the mind-body problem. This is brought out by a slightly different kind of case.

Assume we perceive that an object, a, is black, that another object, b, is green, and that a is larger than b. If properly prepared, we see or apprehend that the two states of affairs, a's being black and b's being green, have the same logical form and that that form differs from the logical form of the state of affairs of a's being larger than b. If there are states of affairs or facts, there are not only the constituent terms and attributes or relations, but the logical forms of such facts as well. Such forms are clearly abstract objects, as that phrase is used. *How* do we manage to apprehend them? While it is not clear either that there is a real question or, if there is, an answer, it is clear that we do know that we apprehend them — and that we apprehend that the forms of the monadic and relational facts differ. The mystery, if such there be, is not about whether there are logical forms. One knows what I am speaking of, even if one doubts that a viable philosophical theory will include such items in its ontological inventory. Whatever mystery there is reduces to our ability to have general ideas — to grasp universals, whether they be qualities like *green*, relations like *larger than* and *diverse from* or logical forms like ϕx and Πxy. The simple fact is that we can do so since we obviously do do so. We need not revert to the nominalistic empiricism, which led to imagist conceptions of thought in the seventeenth, eighteenth and nineteenth centuries and to physicalism and causal theories of reference in the twentieth.

Logical forms are a variety of abstract object, but they appear to highlight a different aspect of the problem that is involved in the present issue. The existence of colors and pitches are hard to deny, though, since many do, we find nominalists and "deflationists" speaking of predicates, i.e. words, being satisfied by objects. There is an obvious similarity between the rejection of abstract entities due to their purported lack of "causal power" and the contemporary appeal of the materialist rejection of mental and physical entities. Physicalists avoid the traditional mind-body problem, as do Berkeleyian style idealists and Humean phenomenalists. But the latter are rare in contemporary philosophy, while the former are numerous. If one commonsensically acknowledges both the mental and the physical, how is the problem of their causal interaction any different from that supposedly posed by universal properties or sets interacting with our cognitive states, as objects of such intentional states (assuming that such interaction is causal in some sense)? It is clear that Benacerraf is thinking in terms of it being a necessary condition for causal interaction that the cause and the effect be in space and time. Thus he speaks of space-time worms (Russell's physical space-time lines):

> It is claimed that X knows that p. We think that X could not know that p. What reasons can we offer in support of our view? If we are satisfied that X has normal inferential powers, that p is indeed true, etc., we are often thrown back on arguing that X could not have come into possession of the relevant evidence or reasons: that X's four-dimensional space-time worm does not make the necessary (causal) contact with the grounds of the truth of the proposition for X to be in possession of evidence adequate to support the inference (if an inference was relevant). The proposition p places restrictions on what the world can be like [1997, 23].

But then it is also clear that the purported problem posed by abstract entities — including numbers — is the same as the supposed problem about mental entities. Thus it would appear that an unstated presupposition is involved, besides the two acknowledged, and quite questionable, principles — a commitment to physicalism. All of those are assumed in addition to presupposing extreme nominalism — that predicates do not refer to properties but are "satisfied" or "true of" or "refer to" particular objects — the objects that would be said to have the properties if one recognized the latter.

Noting the various interpretations of the (P-D) postulates, Russell, like numerous others to follow, took any satisfactory interpretation — one that made use of only the logical vocabulary — to suffice, but he sought one that was the "simplest." The familiar criticisms that there were alternative interpretations were thus acknowledged early on and rejected. Another variant of that criticism later emerged. For example, supposedly, on Russell's view, the class of all classes of three element classes (of individuals), which is dependent on what individuals there are, is the natural number 3. Hence, if one of the individuals, say Whitehead, did not exist, then the number three would be different in a "possible world" which did not contain Whitehead. But, it is argued, numbers are not such "contingent" entities and the number 3 should be the same object in all possible worlds. This type of criticism goes back to the early arguments about the purported "contingency" or factual nature of "existence" claims in *Principia*, especially the dispute over the axiom of reducibility. But the particular argument, regarding specific numbers, reflects a failure to understand Russell's form of "logicism."

Before considering that, we may note that no such questions arise about the set-theoretic versions employing either

$$\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \ldots$$

or

$$\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\} \ldots$$

in order to construe the numbers as sets. But, of more historical interest is that the line of criticism based on logicism taking numbers to be classes that only contingently exist (being contingent on the existence of the elements) cannot be raised about the no-class account in the first edition of *Principia*. As *Principia* class abstracts (expressions) are defined signs — defined in context, such signs, along with definite descriptions, only apparently represent anything, and hence in *Principia* (first edition) classes are not really represented nor taken as existents — not as individuals, nor properties nor functions. The logicism of the first edition of *Principia* takes arithmetical statements in terms of statements with class expressions, where such expressions, in turn, are eliminable in terms of quantified expressions containing only logical signs — variables, connectives, quantifiers (and scope indicating devices). Thus, as we have noted, Russell often speaks of classes (and class signs) as "incomplete symbols," as he speaks of definite descriptions and what is purportedly described. (That the domains over which the quantifiers range are taken in terms of an original sub-structure of individuals, "universals" and atomic facts is a matter we can not go into here. Though it can be noted that the latter do not play a further role in the development of the system as the predicate variables of the formal schema of *Principia* are propositional function expressions and sentential expressions are thought of in connection with "propositions" — which, as the second edition specifies, have "particulars" and "universals" as, respectively, "constituents" and "components.")

The historical inaccuracy of criticisms based on a purported "contingency" aside, the line of argument itself is hardly cogent. What logicism involves, with respect to the question of "What are numbers?", is the construal of statements like (a) "The number of Πs is three" ("There are three Πs.") and (b) "$2 + 3 = 5$" in terms of statements along the lines of: (a) $(\exists x, y, z)[(x \text{ is } \Pi \& y \text{ is } \Pi \& z \text{ is } \Pi) \& (x \neq y \& x \neq z \& y \neq z) \& (\forall w)(w \text{ is } \Pi \leftrightarrow (w = x \vee w = y \vee w = z))]$ (b) The sum-class of a two-element class and a three-element class is a five-element class, where a class, ß, is a two-element class if and only if $(\exists x, y)[(x \varepsilon \beta \& y \varepsilon \beta \& x \neq y \& (\forall z)(z \varepsilon \beta \leftrightarrow (z = x \vee z = y))]$, etc.[18] Such statements neither make claims as to what classes are actually two-element classes, etc., nor as to whether there are $\prod$s. The formulae occur in a formal schema, none of whose primitive terms are non-logical terms as no non-logical terms occur in the calculus of *Principia*. One must not be misled by the use of "English commentary" that accompanies the schema, including its use in "primitive propositions" — as Wittgenstein noted in a well known and pointed comment. (For example 9. 13 cited below, though we should keep in mind that it is numbered while being used to express an inference rule.) It should also be noted that class abstract expressions, being "incomplete symbols" like definite descriptions, require us to take account of their "construction." For example, consider a statement like

---

[18] See [Carnap, 1937, p. 143 ff.]

(a) 2=df. the class of two-element classes.

It is of the pattern:

(b) 2= df. $\hat{u}\theta u$.

(b) allows the replacement of the numeral sign by the class sign, which, as it occurs in some context, say '$f\{\hat{u}\theta u\}$', is to be replaced according to the contextual definition for a class abstract (20.01 cited above) in *Prinicipia Mathematica*, as it is simply elliptical for:

$$(\exists\phi)\colon \phi!x. \leftrightarrow_x \theta x : f\{\phi!u\}$$

Not ony do expressions for numbers disappear, but so do the expressions for classes, and questions about "*what* numbers *are*" are thus removed as well, as are questions about "*what* the King of France (in 2005) *is*." Possible worlds with or without certain individuals are totally irrelevant to the matters involved.

Russell discussed the question of the relevance of empirical matters to arithmetic in *The Problems of Philosophy* and indicated that arithmetical propositions were "general" [1954, 79]. There he reiterated the simple point that, in specifying the "meaning" of numerals, like '2' and '4' one need make use of contexts that only contain variable expressions and quantifiers. Moreover, even in the second edition of *Principia*, where it is suggested that to recognize classes avoids the need for the axiom of reducibility, the particular domains of application are irrelevant to the logical system. As Whitehead and Russell put matters in the first edition, "…*objects* which are neither propositions nor functions. Such objects we shall call *individuals*…" [1950, 51] and in the second edition, "An "individual" is anything that can be the subject of an atomic proposition" [1950, xix]. Thus, it makes no difference what the individuals are taken to be, and hence all one can object to is that the logical system can be applied to different domains. But this is true of any standard system of logic. Though, perhaps the point should be stressed that it is *any* domain (other than the empty domain, if that is a domain). There is also a question about infinite domains and the simple existential "presupposition" that the domain is not empty. Thus the present purported issue regarding contingency can be seen as bringing up an old and familiar issue.

6 STRUCTURALIST ALTERNATIVES TO LOGICISM

Wittgenstein attacked Russell's logicism in another way in the *Tractatus*. He took what amounts to a version of the Peano postulates to provide the common logical form that constituted elementary arithmetic, and as something that could have various applications, in counting objects, for example. Thus, he saw an analogy between such applications and the use of logically valid argument patterns, such as *modus ponens*, that can be applied to arguments with different statements as content. This led him to speak of arithmetic as *a logic*, as opposed to the logicist construal of Russell and Frege, and set forth a form of structuralism in the *Tractatus*.[19] This prepared the ground for what was to come in his later focus on rule following and language — focusing on, in a manner of speaking, the individual "logics" of various kinds of discourse (about colors, ethics, mind, etc., etc.). It is worth noting that, if one accepts complex (structural) properties, one can take the Peano postulates to represent a complex property that is instantiated by any progression and hence can be taken to be the ontological ground of elementary arithmetic. Wittgenstein might even have had something like this vaguely in mind in the *Tractatus*. If he did, his structuralism was ambivalent. The same ambivalence marks Lindström's rejection of logicism and attempt to develop a structuralist variant that does justice to the Russell-Frege requirement that an adequate logical foundation for arithmetic would take the primitive arithmetical signs in (P-D) to have a definite meaning. For Lindström the task is to provide an adequate subject matter for arithmetic. This he seeks to do by claiming to "visualize" or "imagine" an ω-sequence that provides such a subject matter.

---

[19] This is the view suggested, cumulatively, in 4. 1273 and the 6's.

Though Russell recognized logical forms and claimed that we directly apprehend such forms he did not face the problem of justifying that what is apprehended provides the ground of logical truth. Sometimes what he took to be a logical truth was simply what was an axiom or theorem of *Principia*. At other times he held pure mathematics to be composed of logical truths in that mathematical statements expressed logical "implications."[20] And sometimes Russell seemed to take logical truths to be "ontologically grounded" or "made true" by the existence of logical forms and relations among them (which were further complex forms, not "facts"). Russell's speaking of the direct apprehension of such logical forms seems a variant of Gödel's later talking of intuition and perception.

Lindström echoes the appeal to direct apprehension in advocating his "quasi-realism":

> And when this principle is pointed out to us, we can immediately see that it is correct. ...
>
> Now what happens when you "see" that the LNP is correct? As we have seen it is not plausible to say that you are examining an abstract structure. But the answer, or at least one possible answer, is really quite obvious (and simple-minded, almost trivial: You visualize (picture to yourself, form a mental image of)... an ω-sequence. You then examine the imagined sequence ... and find that the LNP [least number principle] is, indeed, true [2000, 126].)

Like Dedekind, he thus provides a subject matter, but by seeing, imagining and visualizing, not by "abstraction," and not by simply examining an abstract structure. The problem is then about his notions of *visualizing* and *seeing*, which are not used synonymously, as well as about *the* sequence being imagined. He apparently sees that the LNP is correct in something like G. E. Moore's sense of "immediately knowing" it to be true, while his use of "visualize" seems to paraphrase talking of "imagining" or "having an image of." We will return to that below. One motive for what Lindström does stems from his taking the so-called *problem of access* seriously. Based, as that purported problem is, on questionable causal theories of reference and supposed puzzles generated by abstraction and our possession of *general ideas*, the problem of access is a classic Carnapian pseudo-problem. Lindström notes in his paper, as Benacerraf does in his, that the problem stems from our supposed ability to have “abstract ideas.” (1996, 14-30) That, as both also mention, led Plato to his version of the doctrine of innate ideas — the Platonic doctrine of “recollection.”

---

[20] It helps to recall that the opening sentence of Chapter I of *The Principles of Mathematics* begins: "Pure Mathematics is the class of all propositions of the form "p implies q," ... [1956, 3].

To avoid supposedly problematic entities to which we have no causal access, Lindström takes the *visualization* of an omega sequence, and, subsequently, of the “branching tree” of the subsets of the elements of that sequence, to provide him with a simple means of access to such elements and sets (though not, as he observes, to all of them). This is another motive behind his quasi-realism [2000, 126-127, 136-137]. Noting that philosophical analyses have little connection with what actually takes place in mathematical reasoning and proof, Lindström claims that an adequate subject matter for first order arithmetic is furnished by our ability to visualize an omega sequence *without* claiming that there really is such a sequence. This ability to “visualize” without ontological commitment to “entities” is responsible for the phrase *quasi-realism*, since, such a form of realism holds that “there is a subject matter” without making any claims about the ontological status of that subject matter — whether there really exists such an omega sequence, and its constituents, that provides the subject matter for first order arithmetic. Hence it is merely “quasi-real,” but its being so suffices to resolve the access problem, since our visualizing such a sequence gives us access to it, while supposedly allowing us to avoid Gödel’s Platonism and Russell’s realism about logical forms (and apparently about classes in the 2^{nd} edition of *Principia*), and other forms of both realism and logicism. To help us see what he means, Lindström likens the visualization of such a sequence to visualizing an unreal object, like a unicorn. The latter, too, would provide a subject matter for a thought or a description. But an obvious and familiar question arises about the truth of claims like “The unicorn I am imagining has only one horn” (which Lindström takes to be clearly true) and “The golden mountain is golden.” The same problems face Lindström’s quasi-realism. That can be seen as pointing to one sort of problem that led Russell to his logicism — the need to take the Peano postulates and the familiar arithmetical truths as true — not merely “true of” some imagined object or possible structure. Lindström also claims that the visualization of the familiar branching tree will serve to provide a “quasi-real” subject matter for set theory, in so far as it is limited to the results of the power set axiom as confined to the sets of elements of the original omega sequence. Beyond that his powers of visualization do not apply.

Such reasoning is not convincing. What it comes down to is the claim that we deal with a certain structure that we assume to be the structure involved in certain “visualizations.” But if we take the term “visualize” literally, and not as a variant of “conceive of” or “imagine,” the visual objects are simply the components of the image of a tree branching, with *the thought* that it goes on ever branching, in the one case, and of a visual sequence of elements, which is thought to indefinitely proceed off into the distance, in the other. If that is not what such theories are about, how do we get to the subject matter of arithmetic and a portion of set theory from such imaging or imagining?

Taking there to be a fundamental difference between the case of a unicorn and that of the omega sequence, Lindström argues that certain problems that seem to arise in the case of the unicorn do not arise about the sequence. We might not be able to ask about the weight of the imagined unicorn, for example, since our imagining a unicorn need not determine it to be of any particular weight. This may recall Meinong's views about imagining and thinking of "incomplete" objects and their connection to our thoughts of non-existent objects [1907, 121; 1915, 171-73]. For once we imagine an omega sequence, any question relevant to its being such a sequence is answered — since any omega sequence will do. In Meinong's sense, such a sequence is not an "incomplete" object, but, rather, follows Dedekind's structuralism. Moreover, one might say the same sort of thing about the unicorn — any question relevant to its being a case of imagining a unicorn, such as its having a horn, can be answered. All Lindström is pointing out, in his example, is that if there were to be a real unicorn it would have to have weight — while there is no property that a real, as opposed to a quasi-real, omega sequence would have to have — since the imagined omega sequence is imagined as having all the properties necessary for it to be a "real" omega sequence. But such matters do not affect what is at issue. One might say that to visualize a unicorn is to visualize a horse with one horn — as that is what is essential to its being a unicorn, and why visualizing a cow with one horn in the middle of its head would not be to visualize a unicorn. But there is clearly not a subject that I speak about when I say "The unicorn I visualize has one horn" as opposed to "The image I now have (visualize) is an image of such and such a kind." For there is no unicorn that there is an "image of," though there is an image of a certain kind. The same is clearly true of the supposedly visualized omega sequence. But images are not the subject matter of arithmetic — nor, for that matter, is what "is visualized" in the sense in which a unicorn, and not an image of a unicorn, is what is supposedly visualized (imagined).

The obvious questions are not about the difference between unicorns and omega sequences but about what is really visualized in such cases and about the connection between what is actually visualized and what Lindström speaks of as the subject matter of first order arithmetic and the portion of set theory he deals with. To put it another way, there is a question as to whether what he does really differs from Dedekind's focusing only on the *distinguishing* of the elements of the system based on their place in the sequence. Here we raise questions of what some would call a phenomenological nature. One aspect of the problem concerns whether the visual object — the image — however vague, is *of* an object. We may assume that there is an image, as a visualized object, since we need not be materialists. One question then concerns the connection, if any, between a thought or case of imagining and a non-existent object; another is a question about the connection, if any, between the image, which we take to exist, and a mythological creature, which does not. The issues posed by quasi-realism, as an alternative to logicism, an alternative that, like logicism, provides a subject matter for arithmetic (other than signs), and thereby is not a variant of *formalism*, *really* concern the second question — a question that we would raise by talking about the apprehended image being a unicorn-image. Is the subject matter of arithmetic composed of images? This is hardly a viable view. Are numbers visualized? What would it mean to say they are? So just what is visualized? What are the elements of the omega sequence? It seems clear that if Lindström is visualizing images, as he visualizes a unicorn-image, he is taking a sequence of images (imagined as *unending*) as the subject matter. But he unequivocally denies that that is what he is visualizing [2000, 126-27]. Thus in talking about visualizing he is talking about intending or apprehending or perceiving — in the manner of Russell speaking of apprehending logical forms and

Gödel of abstract objects.

Lindström's quasi-realism combines a variant of structuralism, which is reflected by his use of "quasi," where the objects structured are supposedly "visualized," and hence not merely structured. Thus such objects provide a subject matter and not just a structure, which accounts for the use of "realism" in "quasi-realism." Moreover, since the omega sequence is merely imagined or visualized, and we can visualize the sequence without making hypothetical existential claims, his notion of "visualization" allows him to provide a subject matter, avoiding formalism, and yet not make ontological commitments that go beyond the common sense idea that arithmetic has a subject matter.

7 LOGICISM AND THE THEORY OF TYPES

Russell considered the paradox that bears his name for years and had noted in 1903, as he stated in the closing sentence of a passage quoted earlier from *The Principles*: "And the contradiction discussed in Chapter x. proves that something is amiss, but what this is I have hitherto failed to discover." He later cryptically noted in *My Philosophical Development* and in the first volume of the autobiography that his theory of descriptions led him to the solution: "It was clear to me that I could not get on without solving the contradictions, and I was determined that no difficulty should turn me aside from the completion of *Principia Mathematica* ... in the spring of 1905 ... I discovered my theory of descriptions, which was the first step towards overcoming the difficulties which had baffled me for so long" [1967, 243]. He did not explain but what he very likely meant was that expressions like "the property had by all properties that are not had by themselves" and "the class of all classes that are not members of themselves" are not to be taken as referential expressions but as contextually defined signs — incomplete symbols — that cannot be employed in purported instantiations to yield a paradox. Thus, with "R" for the Russell property (function) we cannot have the sequence:

i. $(\forall\psi)(R\psi \leftrightarrow \psi\psi)$

ii. $\mathrm{RR} \leftrightarrow \neg\,\mathrm{RR}$

as the paradoxical statement (ii) is not a legitimate instantiation from (i). Rather, the pattern is like the illegitimate sequence:

i. $(\forall x)(x = x)$

ii. the king of France in 1905 = the king of France in 1905.

iii. $(\exists x)(x =$ the king of France in 1905).

Both cases violate the restriction on the use of definite descriptions in such contexts of instantiation, since the additional existential premise — there is such a property (function, class) or there is such a king — is required. Since there is neither such a king, as a matter of fact, nor such a property, as it leads to paradox, the inference can be blocked. But while he now had the reason why the inference didn't hold he still sought a systematic banning of the paradoxical properties (functions) and classes. This was provided by a theory of types prompted by his "vicious circle principle." Yet the theory of types would furnish futher objections to the logicist program.

Aside from the motives for rejecting logicism that have led some to structuralism and formalism, others to finitism and intuitionism, and Lindström to quasi-realism, the purported need to make claims that are not, in any reasonable sense, logical truths has become a standard obstacle to the acceptance of the logicist reduction of arithmetic to logic. Russell was quite aware of the problematic status of three statements the *Principia* system required. To yield the appropriate statements the domain of individuals must be infinite so that for no inductive cardinal **n**, **n** = **n** + 1 (no inductive cardinal is its own successor).[21] In addition what Russell called the "multiplicative axiom," (*Principia*, vol. I, 537), a variant of what became more familiarly known as "choice" and now a familiar part of set theory, "can be enunciated, but not proved, in terms of logic" [1953, 117]. Finally, Russell took the *Principia* system's ramified theory of types to require his admittedly problematic axiom of reducibility that would guarantee there being an "equivalent" predicative function for a function of any order (of a type). Questions about infinity and choice point to issues about the relation of arithmetic to set theory and of the latter to logic: Just what is logic? First order predicate logic? Second order predicate logic? Set theory? What are logical truths?[22] And, questions about the status of existence statements involving "infinitely many" and a selection class, along with the more mundane "$(\exists x)(x = x)$," have led to questions regarding the logical status of "existence" theorems generally: Is it a logical truth that something exists — that there is "something" rather than "nothing"? Similarly, existence assumptions have led some to develop logical systems without such assumptions — some modeled on Russell's restrictions on instantiation to definite descriptions. Since an existential claim follows from the "law" of identity, "$(\forall x)(x = x)$," in standard logical systems, this has helped fuel the dispute about "free logics" — free from existence assumptions — and inferences "about" non-existent objects. In connection with existence assumptions, it was stated in *Principia* that: "Our primitive propositions do not require the existence of more than one individual" [1950, 216].[23] But, given the familiarity of choice and infinity, it was the axiom of reducibility that understandably led to the strongest objections to Russell's logicism and to a suggested change in the second edition of *Principia*. Nevertheless, Russell's treatment of *reducibility*, *choice* and *infinity* clearly indicates that he took none of them to be logical truths.

Russell's ramified theory of types was designed to reject impredicative functions, in the sense of purported functions like $(\lambda x)(\forall f)fx$ and $(\lambda x)(\exists f)fx$, as well as R — or $(\lambda\phi)\neg\phi\phi$ — and an unrestricted truth predicate.[24] The restrictions imposed by the theory led to the introduction of the reducibility axiom to overcome them. One simple difficulty is seen in the case of their definition of identity, since the quantifier must be restricted and cannot cover all "orders" of a type:

---

[21] That is how Whitehead and Russell expressed the assumption that they called the "axiom of infinity." For the conditional manner in which they treated such an "axiom" see *Principia*, vol. II, 201-209, 260-269.

[22] Current usage of "first order" and "higher order" for systems of logic is not the sense of "order" involved in talking of orders within a type in ramified type theory.

[23] In a footnote in *Introduction to Mathematical Philosophy* Russell noted: "The primitive propositions in *Principia Mathematica* are such as to allow the inference that at least one individual exists. But I now view this as a defect in logical purity" [1953, 203].

[24] I will use the lamda notation for functions in place of Russell's *Principia* cap, ˆ, notation in function signs. Russell sometimes takes a "predicative" function (from the term "predicate") to be a first order func-

$$x = y. =: (\phi) : \phi!x \leftrightarrow .\phi!y \qquad \text{Df}$$

Others are posed by mathematical constructions (such as their definition of "natural number," construction of the real numbers) or, as they put it, with the need in mathematics for statements "which will usually be equivalent to what we have in mind when we (inaccurately) speak of 'all properties of x"' [1950, 166].[25] But as the axiom raised an obvious problem regarding its status as a purported truth of logic, it cast further doubts on the purported logistic reduction of standard arithmetic to logic. Thus Russell sometimes held to the view that logical truths were conditionals. It was as if logicism was weakened to the claim that from the primitive propositions of *Principia Mathematica* "the" conditional — (infinity & choice & reducibility) → (P-D) — was derivable.

Ramsey's well-known critique of the reducibility axiom, purporting, among other things, to prove it was not a logical truth, involved his problematic view that the quantifiers were construable in terms of infinite conjunctive and disjunctive functions. (1931, 9-20, 56-61)[26] Issues about Ramsey's view aside, it has an ironic twist. For the reducibility axiom

---

tion of individuals, where the expression for it contains no quantifiers or only quantifiers over individuals– $\phi x, \psi xy, (\exists x)\psi xy, \ldots$ [1950, 51-52]. Sometimes a predicative function is characterized as a "function of one variable which is of the order next above that of its argument" and of more than one variable if there is one for which it becomes predicative when values are assigned to the others [1971, 78]. Sometimes it is characterized as one of the next order above that of its argument(s) — when it is "of the lowest order compatible with its having that argument" [1950, 53]. He also sometimes speaks of properties as predicates and as functions. As we need not explore such matters, I will simply speak of functions and not try to specify just what Russell's ramified theory of types is (different versions were suggested at different times, and there are grounds for holding that *Principia Mathematica* does not contain one uniform version). Though not meant to be an accurate account of Russell's theory, one fairly standard account, followed, for example, in Copi [1971] will do for getting the basic ideas. On that account, we may think of functions arranged in a hierarchy of types, so that functions of type **t** may not take arguments of types higher than **t**-1 (whether they can only take arguments of type **t**-1 depends on how restrictive the variant of the ramified type theory is). Stopping at this point yields the simple theory of types. A ramified theory divides the functions within each function type into orders, such that a quantifier of order **o** ranges over functions of order **o**. A predicate abstract containing a predicate quantifier sign of order **o**, as the quantifier of highest order occurring in the abstract, is then a sign for a function of at least order **o**+1.

[25] One oddity of the Principia theory of types is the recurrence of the natural numbers at each type from the second type (for classes or functions) "on up." In short one gets an image of the (P-D) axioms taking numbers as different classes of classes at each such level. Thus, that such a number is not, in a sense, one definite object is a part of the Russell-Whitehead construction. The same sort of thing occurs in the case of "=."

[26] Such a view, oddly, given Russell's well known arguments about general facts, can be seen to be suggested in *Principia* (where he notes that only finite combinations or disjunctions can be taken extensionally or "by enumeration"):

> The axiom of reducibility is equivalent to the assumption that "any combination or disjunction of predicates is equivalent to a single predicate," i. e. if we assert that x has all the predicates that satisfy a function $f(\phi!\hat{u})$, there is some one predicate which x will have. . . and similarly if x has some one of the predicates that satisfy a function $f(\phi!\hat{u})$. . . . hence we shall be able to get from any non-predicative function to a formally equivalent predicative one [1950, 58-59].

This is relevant to a complicated question of interpretation of the *Principia* system that is related to Russell's Fregean style view that "predicates" can only function "predicatively" and not as terms. This view he took to be reinforced by the need to deal with the logical paradoxes as well as, like Frege, taking it to resolve the can be dispensed with by employing such infinite disjunctive and conjunctive functions and taking "=" to be an additional primitive, logical sign (with standard axioms). Since, for each individual a, there is a function $(\lambda x)(x{=}a)$, there will be a function, a disjunction of such identity functions, for every subset of the domain. Such a function would be predicative if identity is primitive. Then, trivially, for every function of individuals, there is an equivalent predicative function of individuals. Russell, defining the identity sign in *Principia*, cannot make use of such a pattern. But, given the reducibility axiom, he can, and does, confine the definition of identity to predicative functions.

Unlike Ramsey, who argued that the axiom of reducibility was not a logical truth, Wittgenstein argued that the axiom could not be sensibly stated. In a letter to Russell, Wittgenstein wrote:

> Your axiom of reducibility is
>
> $$:(\exists f):\phi x \leftrightarrow_x .f!x;$$
>
> now is this not all nonsense as this proposition has only then a meaning if we turn the ϕ into an apparent variable .... The axiom as you have put it is only a schema and the real Pp ought to be
>
> $$:(\phi):(\exists f):\phi x \leftrightarrow_x .f!x.$$
>
> and where would be the use of that? [1969, 122]

Wittgenstein's claim is that the axiom cannot be stated without an unrestricted quantifier, which violates the restrictions of the ramified theory of types. What one can state are indefinitely many axioms (or meta-axioms or an axiom schema or statements in a background language or a typically ambiguous axiom) for various orders of functions. It is as if one were to state, in a background schema, that there is an axiom for every order of function of the system. But this background statement involves a quantified expression not governed by "the axiom" itself. Wittgenstein's claim is that any statement of the axiom must explicitly or implicitly violate the point of the ramified theory of types. Yet, without such an axiom, or something equivalent to it, the problems Russell and Whitehead noted about the ramified theory of types remain.

Wittgenstein's criticism overlooks a distinction that lies behind Russell's way of stating the axiom. The distinction was based on the supposed difference between "all" and "any." As Russell put it in 1908:

> If ϕx is a propositional function, we will denote by '$(x).\phi x$' the proposition 'ϕx is always true' .... Then the distinction between the assertion of all values and the assertion of any is the distinction between (1) asserting $(x)\,.\phi x$ and (2) asserting ϕx where x is undetermined. The latter differs from the former in that it cannot be treated as one determinate proposition .... In the case of such variables as propositions or properties, 'any value' is legitimate, though

---

"Bradley paradox" [1971, 337-338]. The view was one target of Ramsey's paper "Universals" that attacked the universal-particular distinction [Ramsey, 1931, 112-134].

> ‘all values’ is not. Thus we may say: ‘p is true or false, where p is any proposition’, though we cannot say ‘all propositions are true or false’. The reason is that, in the former, we merely affirm an undetermined one of the propositions of the form ‘p is true or false’, whereas in the latter we affirm (if anything) a new proposition different from all the propositions of the form ‘p is true or false’. Thus we may admit ‘any value’ of a variable in cases where ‘all values’ would lead to reflexive fallacies.... [1971, 66-67].

Moreover, in commenting on the axiom of reducibility, as written by Wittgenstein above, Russell explicitly applied his distinction between “all” and “any” to that axiom:

> This is the axiom of reducibility. It states that, given any function ϕx, there is a predicative function $f!x$ such that $f!x$ is always equivalent to ϕx. Note that, since a proposition beginning with ‘$(\exists f)$’ is, by definition, the negation of one beginning with ‘(f)’, the above axiom involves the possibility of considering ‘all predicative functions of x’. If ϕx is *any* function of x, we cannot make propositions beginning with ‘(ϕ)’ or ‘$(\exists\phi)$’ since we cannot consider ‘all functions’, but only ‘*any* function’ ... [1971, 87].

The problem, however, is whether Russell’s distinction between “all” and “any,” and its connection with his notion of an “undetermined value” (and related notions like “ambiguous denotation,” “ambiguous statement” and “statement about an ambiguity”) suffices to make his statement of the axiom viable in the face of Wittgenstein’s criticism. Aside from any concern with Russell’s views of denotation and ambiguous denotation, there is an obvious problem raised by the claim that we can assert an indeterminate proposition, by the use of a free variable , with the expressive power of a universally quantified (apparent) variable — while holding that no determinate proposition is thereby asserted. It is clear that Russell’s distinction between “all” and “any” will not, by itself, do as a reply to Wittgenstein’s criticism. Thus, at one place, Russell admits that some primitive propositions of *Principia* cannot be symbolized in a standard way but must be expressed in words in a meta-language. And, he goes on to suggest the introduction into *Principia* of a new symbolic device to carry the sense of the words used in the meta-language. Thus the symbol ‘[ϕy]’ is suggested in the primitive proposition:

$$: [\phi y] \rightarrow (x).\phi x$$

to symbolize the “mere hypothesis of what is asserted in ‘ϕy’” as the condition involved in holding “If ϕy is true however y may be chosen, then (x). ϕx is true .” (*Principia*, 1950, 132) Nevertheless, Russell’s discussion of the use of a free variable in the axiom of reducibility is not consistent with the primitive proposition of *Principia* that amounts to a version of the rule of universal generalization.

> *(9. 13) In any assertion containing a real variable, this real variable may be turned into an apparent variable of which all possible values are asserted to satisfy the function in question. (*Principia*, 1950, 132)

Of course, one can argue that *(9. 13) does not apply in the case of the axiom of reducibility because the resulting universal generalization is an "illegitimate" statement. But this would merely reveal the specious use of a free variable to state the axiom. What Wittgenstein is claiming is that whether one appeals to an axiom schema or to typically ambiguous axioms or to meta-linguistic statements one makes a devious move by both accepting the axiom and violating its intent. This becomes apparent when, in the second edition of *Principia*, Whitehead and Russell take all "asserted" propositions with free (real) variables to be replaced by propositions with universally quantified variables. (1950, xiii) Moreover, in the second edition, it is suggested that the axiom of reducibility could be dispensed with by assuming the existence of classes:

> It should be observed, in the first place, that if we assume the existence of classes, the axiom of reducibility can be proved. For, in that case, given any function ϕz of whatever order, there is a class α consisting of just those objects which satisfy ϕz. Hence, "ϕx" is equivalent to "x belongs to α." But "x belongs to α" is a statement containing no apparent variable, and is therefore a predicative function of x. Hence, if we assume the existence of classes, the axiom of reducibility becomes unnecessary [1950, 58].[27]

On the surface Russell's argument is puzzling. For it can be seen to suggest that if we have a function, say $(\lambda x)(\forall f)fx$, and accept the existence of classes, we can *refer to* the class of objects that satisfy that function by a sign of the system, "α" in the above quotation, and form the predicate "$(\lambda x)(x\varepsilon\alpha)$." This predicate can then be taken to represent a predicative function that is equivalent to $(\lambda x)(\forall f)fx$. Russell seems to be suggesting that we can define the expression "α" in terms of the expression "$(\lambda x)(\forall f)fx$" and then proceed to use 'α' instead of using a class abstract. For, given the existence of classes, we supposedly have the class for the function $(\lambda x)(\forall f)fx$. And one stipulates that α is that class or that "α" refers to it. Since "α" is not an expression containing a predicate quantifier, a predicative function is represented by "$(\lambda x)(x\varepsilon\alpha)$." What Russell appears to be doing is claiming that we can introduce a sign, say "α," to stand for any class we specify in terms of a class abstract, and then refer to the class by means of that sign, "α," to form a function sign. If he were doing that, then he need not resort to classes to turn the trick. Recognizing the function $(\lambda x)(\forall f)fx$, all he need do is call it "$(\lambda x)\phi!x$" and, as this latter expression contains "no apparent variable," hold that the function it represents is elementary. This is not what Russell is doing. For, according to *Principia*, we take "$(\iota\alpha)(f\alpha)$" as "the class which satisfies the function f" and "it is to be remembered that 'α' stands for '$\hat{u}(\phi u)$,' and that '$f\alpha$' therefore stands for '$f\{\hat{u}(\phi u)\}$'" [1950, 190]. Thus, he is simply claiming that where there is a class, α , specified by a function that is not predicative, then there also is a predicative function, *is an element of* α, that is trivially equivalent to the original function used to specify the class.[28] Given that there is a class α, then, by the way Russell has developed the *Principia* system, there is the function $(\lambda x)x\varepsilon\alpha$. The question that arises is whether the sign "α," as a class sign, is to be considered as defined in context, and thus the function expression “(λx) x $\varepsilon\alpha$” is understood in terms of the pattern:

---

[27] This had already been set out in Russell's 1908 paper "Mathematical Logic as Based on the Theory of Types," where the axiom of reducibility was alternatively called the "axiom of classes" [1971, 81-82].

[28] For a different interpretation of what Russell is doing see [Kneale, 1962, 663-664].

$$(\exists\phi){:}\ \phi!x.\ \leftrightarrow_x .\psi x{:}f\{\phi!u\}$$

and, hence:

$$(\exists\phi){:}\ \phi!x.\ \leftrightarrow_x .\psi x{:}x\varepsilon!\phi\hat{u}\}.$$

The assumption that there are classes is what allows for the use of ‘$x\,\varepsilon!(\phi!\hat{u})$’ as expressing a predicative context. Such a use of the sign ‘ε‘, where it does not occur with a class abstract (that must be replaced via its contextual definition), was provided for in

$$20.02 \quad x\varepsilon!(\phi\hat{u}).=.\phi!x \qquad \mathrm{Df}$$

which, along with 20.01 (cited earlier), “results” in

$$x\varepsilon\hat{u}(\psi u)\}.\leftrightarrow{:}(\exists\phi){:}\ \psi y.\ \leftrightarrow_y .\phi!y{:}\phi!x$$

and

$$x\varepsilon\hat{u}(\psi u).\leftrightarrow .\psi x,$$

the latter “by the help of” the axiom of reducibility [1950, 188] It was that “help” that the existence of classes now made unnecessary. Thus Russell does not merely introduce another sign for the same function, as in the case of ‘$(\lambda x)F_1x$’ and ‘$(\lambda x)(\forall f)fx$’. Rather, he takes a class to provide a function that is predicative — along with the assumption that classes exist. The assumption that classes exist then becomes fundamental to the variant of logicism suggested in the introduction to the second edition of *Principia* and raises an additional question about the purported “logical nature” of the assumptions required to attempt to carry out the logicist’s program.

Russell’s revised version of logicism, recognizing classes, reiterates the general question as to whether logic can be divorced from ontological presuppositions — a question that can be raised in connection with the logical connectives (negative facts, non-existent states of affairs, objects), quantifiers (generality, existence) as well as the “domains” over which the variables range. Thus one may argue that “logic,” aside from any considerations about arithmetic and logicism, already involves existential presuppositions. There are the additional questions about the nature of logic that focus on acceptable proof procedures. Thus “constructionist” and “finitist” themes, associated with “intuitionism” in the philosophy of mathematics are perhaps better taken as questions about the logical foundations themselves. Increasingly, issues about logic itself have come to the front: Is first order predicate logic to be taken as “logic”? Is (standard) set theory logic? Are “free,” “paraconsistent,” “intuitionist,” etc. *systems* logics? The question “What is logic?” is more complex today. Yet Russell had recognized, from the first decades of the 20^{th} century, the “pragmatic” nature of the problematic assumptions he needed to carry out the logistic program. This pragmatic aspect can be taken to have culminated in Carnap’s holding that such pragmatic justifications were all that could be expected, since taking there to be some further kind of “proof” or establishing of the correctness of a formal system would involve failing to recognize that questions leading to such attempts were “external” and hence meaningless. In that way Carnap and the positivists, viably or not, had followed Russell’s early lead in suggesting that logicism embodied a “proposal” to be judged by its fruitfulness.

BIBLIOGRAPHY

[Benacerraf, 1965] P. Benacerraf. What numbers could not ne, *Philosophical Review*, 74, 1965.

[Benacerraf, 1996] P. Benacerraf. Mathematical truth, in *The Philosophy of Mathematics*, (ed. W. D. Hart), Oxford, 1996.

[Bernays, 1964] P. Bernays. On Platonism in mathematics, in *Philosophy of Mathematics*, (ed. P. Benacerraf & H. Putnam), Englewood, 1964.

[Bradley, 1944] F. H. Bradley. *Essays on Truth and Reality*, Oxford, 1944.

[Carnap, 1937] R. Carnap. *The Logical Syntax of Language*, London, 1937.

[Carnap, 1956] R. Carnap. Empiricism, semantics, and ontology in R. Carnap, *Meaning and Necessity*, Chicago, 1956.

[Carnap, 1959] R. Carnap. *Introduction to Semantics and Formalization of Logic*, Cambridge, Mass, 1959.

[Carnap, 1967] R. Carnap. *The Logical Structure of the World & Pseudo Problems in Philosophy*, (trans. R. George), Berkeley, 1967.

[Copi, 1971] I. Copi. *The Theory of Logical Types*, London, 1971.

[Dedekind, 1888] J. W. R. Dedekind. *Was sind und was sollen die Zahlen*, Brunswick, 1888.

[Dummett, 1973] M. Dummett. *Frege's Philosophy of Language*, New York, 1973.

[Dummett, 1978] M. Dummett. *Truth and Other Enigmas*, Cambridge, Mass, 1991.

[Dummett, 1991] M. Dummett. *Frege. Philosophy of Mathematics*, Cambridge, Mass, 1991.

[Dummett, 2002] M. Dummett. What is mathematics about? in *Philosophy of Mathematics*, (ed. D. Jacquette), Oxford, 2002.

[Evans, 1982] G. Evans. *The Varieties of Reference*, Oxford, 1982

[Frege, 1953] G. Frege. *The Foundations of Arithmetic* (trans. J. Austin), Oxford, 1953

[Frege, 1967] G. Frege. *The Basic Laws of Arithmetic*, (trans. M. Furth), Berkeley, 1967

[Frege, 1968] G. Frege. The thought. A logical inquiry, (trans. A. Quinton) in *Essays on Frege*, ed. E. D. Klemke, University of Illinois Press, Urbana, 1968.

[Frege, 1970] G. Frege. Negation, *Translation from the Philosophical Writings of Gottlob Frege*, (trans. P. Geach and M. Black), Oxford, 1970.

[Frege, 1972] G. Frege. *Conceptual Notation and Other Writings*. (trans. T. W. Bynum), Oxford, 1972.

[Frege, 1975] G. Frege. *Wissenschaftlicher Briefwechsel*, Hamburg. E. D. Klemke, Urbana, 1975.

[Gill, 1990] R. R. Gill. *Deducibility and Decidability*, London, 1990.

[Gödel, 1986] K. Gödel. Is mathematics syntax of language?, *Kurt Gödel. Collected Works* vol. III, (ed. S. Feferman *et. al.*), Oxford, 1986/95.

[Hale and Wright, 2001] B. Hale and C. Wright. *The Reason's Proper Study*, Oxford, 2001.

[Hilbert, 1964] D. Hilbert. On the infinite, in *Philosophy of Mathematics*, (ed. P. Benacerraf & H. Putnam), Englewood, 1964

[Hochberg, 1956] H. Hochberg. Peano, Russell and logicism', *Analysis*, 25, 3, 1956.

[Hochberg, 1994] H. Hochberg. Causal connections, universals and Russell's hypothetico-scientific realism, *The Monist*, 77, 1, 1994.

[Kneale and Kneale, 1962] W. Kneale and M. Kneale. . *The Development of Logic*, Oxford, 1962.

[Lewis, 1991] D. K. Lewis. *Parts of Classes*, Oxford, 1991.

[Lindström, 2000] P. Lindström. Quasi-realism in mathematics," *The Monist*, 83. 1, 2000.

[Moore, 1889] G. E. Moore. The nature of judgment, *Mind*, 30, 1889.

[Peano, 1887] G. Peano. *Arithmetices principia nova methodo exposita* Rome, 1887.

[Ramsey, 1931] F. P. Ramsey. *The Foundations of Mathematics*, (ed. R. Braithwaite), London, 1931.

[Ramsey, 1991] F. P. Ramsey. *Notes on Philosophy, Probability and Mathematics*, (ed. M. C. Galavotti), Naples, 1991.

[Russell, 1906] B. Russell. On the nature of truth, *Proceedings of the Aristotelian Society*, 7, 1906/7.

[Russell, 1952] B. Russell. *Our Knowledge of the External World*, London, 1952.

[Russell, 1953] B. Russell. *Introduction to Mathematical Philosophy*, London, 1953.

[Russell, 1954] B. Russell. *The Problems of Philosophy,* London, 1954

[Russell, 1956] B. Russell. *The Principles of Mathematics*, Cambridge, 1956.

[Russell, 1967] B. Russell. *The Autobiography of Bertrand Russell 1872-1914*, Boston, 1967.

[Russell, 1971] B. Russell. Mathematical logic as based on the theory of types, reprinted in *Logic and Knowledge*, (ed. R. C. Marsh), New York, 1971.

[von Neumann, 1964] J. von Neumann. The formalist foundations of mathematics, in *Philosophy of Mathematics*, (ed. P. Benacerraf & H. Putnam), Englewood, 1964.

[Whitehead and Russell, 1950] A. N. Whitehead and B. A. W. Russell. *Principia Mathematica*, vols. I, II. Cambridge, 1950.

[Wittgenstein, 1963] L. Wittgenstein. *Tractatus Logico-Philosophicus*, (trans. D. Pears & B. McGuinness), London, 1963.

[Wittgenstein, 1969] L. Wittgenstein. *Notebooks. 1914-1916*, (trans. G. Anscombe), New York, 1969.

[Wright, 1983] C. Wright. *Frege's Conception of Numbers as Objects*, Aberdeen, 1983.

CLASSICAL LOGIC'S COMING OF AGE

John W. Dawson, Jr.

1 INTRODUCTION

The displacement of the concepts and methods of Aristotelian logic by modern symbolic and mathematical ones began in the nineteenth century with the works of George Boole, Charles Sanders Peirce, Ernst Schröder, Georg Cantor, Giuseppe Peano and Gottlob Frege. Each of those pioneers contributed key conceptions that have become standard features of modern formal logic. But the integration of their ideas into a conceptual framework that incorporated both semantic and syntactic notions and stressed the interplay between them — one in which *metalogical* questions could be posed and answered — took nearly a century: almost exactly so if we reckon from the publication of Boole's *The Mathematical Analysis of Logic* in 1847 until Leon Henkin's doctoral dissertation of 1947, wherein the method of constants was introduced and used to extend Gödel's completeness theorem to theories formulated in languages of arbitrary cardinality.

Within that century three formative periods in the development of logic may be distinguished: its youth, from 1847 until around 1900; its adolescence, from then until 1928; and its coming of age, during the following two decades.

The first of those periods was distinguished by the recognition, by Boole, Augustus De Morgan, Schröder and others, of the *mathematical character* of logic (in particular, of analogies between the operations of arithmetic and algebra and those of propositional logic); by the elucidation, by Frege and Peirce, of *quantification theory* (the logic of statements involving quantifiers and polyadic predicates); by the development, especially through the work of Schröder and Peano, of a compact *notation* for exhibiting the logical structure of statements; by Cantor's creation of *naive set theory* and the notions of *infinite cardinal and ordinal numbers*; and by the introduction of *formal axiomatizations*, of arithmetic (by Peano), Euclidean and projective geometry (by Moritz Pasch, David Hilbert and Oswald Veblen), and the real numbers (by Hilbert and Edward V. Huntington).

The period of adolescence was marked by polemical disputes and foundational uncertainty, engendered both by paradoxes that had emerged and by the construction of pathological functions in analysis. Highlights of that era included the discovery of *Russell's antinomy*; Ernst Zermelo's *axiomatization of set theory*; controversy over the legitimacy of the *axiom of choice* (sparked in part by such counterintuitive consequences as the Banach–Tarski paradox); disagreement

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

over the *interpretation of quantifiers*; the proof of the *Löwenheim-Skolem Theorem*, which implied that first-order axiomatizations of set theory admit countable models, despite references within those theories to sets that are uncountable; and the development of *competing mathematical philosophies*, by Frege, Bertrand Russell and Albert NorthWhitehead (logicism), L.E.J. Brouwer and Hermann Weyl (intuitionism), and Hilbert (formalism).

The present article focuses on the third period, during which *first-order* logic became the focus of logical investigation, questions of *consistency, completeness* and *categoricity* were addressed and distinguished from one another, the notion of *truth* in formal languages was inductively defined, the fundamental notions of *recursion theory* were formulated, enabling formal proofs of *undefinability* and *undecidability*, and the axiom of choice and Cantor's *continuum hypothesis* were shown to be consistent relative to the other axioms of set theory. The principal figures of that period included Alfred Tarski, Kurt Gödel, Alonzo Church, Alan Turing, Stephen C. Kleene, J. Barkley Rosser, and Emil Post.

2 THE RISE TO ASCENDENCY OF FIRST-ORDER LOGIC

In axiomatizing mathematical theories it is natural to quantify over *sets* or *properties* of the individual objects of mathematical interest. It is not surprising, therefore, that all of the axiomatizations referred to above possessed this *second-order* character. Peano's induction axiom, for example, stated that *any class* of natural numbers that contained 1 and contained the successor of any number it contained must contain all the natural numbers; Hilbert's and Huntington's axiomatizations of geometry and of the real numbers involved second-order axioms of continuity or completeness; and the axiom of separation in Zermelo's first formulation of set theory [Zermelo, 1908] involved quantification over what he called *'definit'* propositional functions.

Subsequently, however, until quite recently, the attention of classical logicians came to be focused almost exclusively on *first-order* logic, in which the quantifiers range only over *individual* variables, representing objects of a domain of interpretation that is left unspecified within the theory. How that came about has been explored in two detailed and incisive studies [Moore, 1988; Goldfarb 1979], upon which the summary account that follows is based.

The system of quantification theory that Frege introduced in his *Begriffsschrift* [Frege, 1879], which formed the basis for his later attempt to derive arithmetic from logic [Frege, 1893], involved quantification over functions. The unrestricted use of function variables caused Frege's system to be compromised by Russell's antinomy, and in an effort to rescue the logicist program from that defect Russell set about developing the theory of types, which he and Whitehead expounded in *Principia Mathematica* [Whitehead and Russell, 1910].

It would have been possible for Whitehead and Russell to restrict attention to variables of lowest type and to quantifiers ranging only over them, but they did not do so. Indeed, as Goldfarb stresses, from the logicist perspective it would not have made *sense* to do so, since logicists viewed logic as concerned with *everything*, not just individuals; its generality was not understood, as it is today, in the sense of "having no subject matter in particular". For the same reason, metalogical questions also lay outside the purview of logicism, because it was impossible, from that point of view, to step outside the universe of *all* things to which logic applied.

In contrast, Peirce's system of quantification theory, developed as part of his "logic of relatives", belonged to the algebraic tradition in logic. It, too, included second-order quantifiers, although in his 1885 paper "On the algebra of logic. A contribution to the philosophy of notation" [Peirce, 1885] Peirce focused primarily on the "first-intentional logic of relatives", involving quantification just over individuals. But Peirce (and much later, Zermelo as well) construed quantifiers as infinitary conjunctions and disjunctions, over a domain of discourse that was *fixed* in advance. In addition, since Peirce's treatment was not axiomatic and did not employ function symbols, it was far from what we understand today as a formal logic.

Leopold Löwenheim, in his paper [1915], presented a version of what later became formulated as the downward Löwenheim-Skolem theorem. He showed that an 'individual expression' (*Zählausdruck*) of the logic within which he worked is satisfiable in a denumerable structure if it is satisfiable at all. Löwenheim distinguished 'individual expressions' from 'relational' ones (*Relativausdrücke*). The former contained only variables for individuals, and were equivalent within his system to finitary expressions. But his logic itself was both infinitary and second-order, and his proof of the theorem made use of a second-order formula. Thus, as Moore points out, though Löwenheim in a sense distinguished first- from second-order *expressions*, he did not work within what we now call first-order *logic*.

It was Thoralf Skolem, in an address to the Fifth Congress of Scandinavian Mathematicians (1922), who first explicitly stated that a generalization of Löwenheim's theorem was valid in the latter framework [Skolem, 1923b].[1] The aim of Skolem's remarks was to place set theory "on the same level as other axiomatic theories", by demonstrating the "relativity of set-theoretic notions". He intended thereby to show that set theory is unsuitable to serve as a basis for the rest of mathematics.

But does the Löwenheim-Skolem theorem show the foundational inadequacy of set theory, or, as Skolem's critics (most notably, Zermelo) maintained, the inadequacy of first-order formalizations? According to a theorem of Per Lindström, proved in the 1960s, first-order logic is the strongest logic that satisfies both the downward Löwenheim-Skolem theorem and the countable compactness theorem (which states that a countable set of sentences is satisfiable if every finite subset is). And though compactness is generally regarded as a desirable property for a logic to have, Lindström's theorem does not entail that first-order logic is the only reasonable logic unless one demands that the Löwenheim-Skolem property should also hold — a position that has been deemed tantamount to "deny[ing] reality to the concept of uncountability" [Wang, 1974, 154].

---

[1] Three years earlier Skolem had given a proof of Löwenheim's theorem that did not employ any second-order formula, but his statement of the result then was in the context of the language now called $L_{\omega_1\omega}$, which allows countably infinite conjunctions of first-order formulas.

Skolem introduced first-order propositions as a way of explicating Zermelo's vague conception of '*definit*' property. His stated justification for doing so was that the notion of a finite expression built up by applying connectives and quantifiers to atomic formulas involving only individual variables was "a completely clear notion"— one which, he claimed, was "sufficiently comprehensive... to carry out all ordinary set-theoretic proofs"[Skolem, 1923b]. Subsequently, Skolem never wavered in his advocacy of first-order logic — a position that Zermelo ridiculed as "Skolemism, the doctrine that *every* mathematical theory... is realizable in a countable model" [Zermelo, 1932].

A version of first-order logic, referred to as "the restricted functional calculus", was also singled out for attention in Hilbert and Ackermann's textbook *Grundzüge der theoretischen Logik* [Hilbert and Ackermann, 1928], a reworking of lectures on logic that Hilbert had given in 1917. Hilbert introduced the restricted functional calculus, incoporating both propositional and first-order predicate logic, in chapter 3, but went on in the next (final) chapter to argue the need to extend it, in order both to express the full intent of the axiom of complete induction and (contrary to Skolem's view) to be able to adequately axiomatize set theory. The logical framework Hilbert ultimately adopted was a version of type theory, built up in stages, beginning with a given, fixed domain of individuals and certain basic predicates on them. He defined a predicate of the first level to be "a logical function of one or several variables that may be built up from the basic predicates" by applying finitely many propositional connectives together with "the operations 'all' and 'there exists' ", the latter "referring only to the original domain of individuals." Second-order theories were then obtained by taking the functions and propositions of the first level as a new domain of individuals, and so on. To avoid the paradoxes, the levels were ramified and the substitution of functional or propositional expressions for functional or propositional variables was restricted to expressions of "the same or lower level".

For Hilbert the retricted functional calculus thus constituted the first step in the construction of a more comprehensive logical framework that subsumed it. Merely as such, the restricted calculus would not have merited special scrutiny. But before going on to his extended system Hilbert directed attention to a pair of metalogical questions concerning the restricted calculus that he declared to be of "fundamental importance" for the study of "all mathematical domains amenable to finite axiomatization": the *decision problems* for validity and satisfiability.

3 SEMANTIC DECIDABILITY AND SEMANTIC COMPLETENESS

Hilbert posed the decision problem for logical validity as the question: "For an arbitrarily given logical expression [of the restricted functional calculus] that contains no individual constants, how can one determine whether the expression, under any substitution whatever [of particular individuals] for the [free] variables occurring in it, does or does not [always] represent a correct assertion?" Similarly, Hilbert posed the decision problem for satisfiability as the question "Whether there is any substitution at all [of particular individuals] for the [free] variables under which the corresponding expression will represent a correct assertion." He noted that the two questions were dual to one another, since a formula will be logically valid if and only if its negation is not satisfiable, and that both problems had been solved for the propositional calculus, where either the method of truth tables employed by Post (*1921*) or that of conversion to conjunctive normal form (introduced by Paul Bernays in his *Habilitationsschrift* of 1918 and described in an earlier chapter of [Hilbert and Ackermann, 1928] can be used to ascertain whether or not a propositional formula is a tautology. To judge from his phrasing of the decision problem for validity (which asks *how*, not *whether*, a determination can be made), Hilbert seems to have expected that the two decision problems would have a positive solution in the case of first-order formulas as well. Toward that end, the last section of chapter 3 of Hilbert and Ackermann's text gave decision procedures for certain special classes of formulas.

Validity and satisfiability, as defined above, are semantic notions. But in his statement of the decision problem, which he declared to be "the principal problem of mathematical logic" [Hilbert and Ackermann, 1928, 77], Hilbert placed no restriction on the *means* that a decision procedure for validity or satisfiability might use. In the context of propositional logic, for example, the decision procedure based on truth tables is normally construed to be a semantic method, though Post apparently did not regard it as such (see the discussion on page 46 of [van Heijenoort and Dreben, 1986], whereas Bernays' procedure based on conversion to normal form is a syntactic one. And if one approaches the decision problem syntactically, one may be led to consider the related question of *semantic completeness*: Is every logically valid formula of the restricted functional calculus formally derivable by means of the axioms and rules of inference of that calculus? Indeed, Hilbert had posed that question even before discussing the decision problem. It was, he said, a "still unsolved problem", for which there was then "only empirical evidence" that a positive solution might be expected [Hilbert and Ackermann, 1928, 68].

In fact, the step most essential for that expected solution had already been carried out, in the paper of Skolem cited above. For there Skolem showed that to every first-order formula F of a certain normal form (to which any first-order expression is reducible) may be associated a countably infinite sequence of quantifier-free formulas, each of which is a finite conjunction of substitution instances of F; and if each formula of that sequence is truth-functionally satisfiable, then F is satisfiable within the domain of natural numbers. In a further paper [Skolem, 1929] Skolem refined his argument and concluded that if the formula F were *not* satisfiable, then not only must one of the formulas of the associated sequence be truth-functionally unsatisfiable, but from it a contradiction would be informally deducible. To establish completeness,[2] it thus remained only to show how to transform such an informal deduction into a formal proof of the negation of F. But Skolem did not do so, and he apparently did not recognize that the chain of reasoning just described would settle the issue. In part, that may have been due to his conflation of semantic and syntactic notions (satisfiability and consistency). More importantly, though, it seems that he simply was not interested in the matter. (For a detailed discussion of Skolem's aims and accomplishments in this sphere, see [Goldfarb, 1979; van Heijenoort and Dreben 1986].)

---

[2] In classical logic, where any formula is logically equivalent to its double negation

It was Kurt Gödel who, in his doctoral dissertation [Gödel, 1929, published in revised form as [Gödel, 1930a], reconstructed Skolem's argument, carried out the final syntactic step, and presented the whole as a solution to the question posed by Hilbert and Ackermann. Mathematically, Gödel went only slightly beyond what Skolem had already done. But he brought to the subject a clarity of vision that had been lacking in all that had gone before; he distinguished clearly between syntactic and semantic concepts; and he was aware of the larger philosophical significance of the result he had proved. For in the introduction to the dissertation (though not in its published revision) he remarked that the assertion of completeness could "easily be seen to be equivalent" to the statement that every consistent set of first-order axioms has a concrete realization — and that, he noted, justified "the usual method for proving consistency", since it guaranteed that "one must either be able to produce a contradiction" from a set of statements or else "prove the[ir] consistency by means of a model." In short, he recognized that he had established a stronger result than Hilbert and Ackermann had asked for: He had shown not just that the logical axioms of their system were adequate for deriving all logically valid statements, but that if nonlogical axioms (such as those of Peano for the natural numbers) were added to them, the statements derivable from the augmented system would coincide with those true in *every* interpretation that satisfied the nonlogical axioms.

In addition, he drew an epistemological conclusion from the fact of completeness. One might think, he said, (with tacit reference to Hilbert's view) that "the existence of notions introduced through an axiom system" is established "outright" once it is demonstrated that no contradiction can be derived from them. But such a belief, he cautioned, "manifestly presupposes" (as Hilbert had explicitly claimed) that "every mathematical problem is solvable."

Indeed, were some statement to be shown undecidable (neither provable nor refutable) within a particular first-order axiomatic framework, it would follow from the completeness theorem that two non-isomorphic realizations of those axioms existed, one satisfying the statement in question and the other satisfying its negation. That would not be a contradiction, but it would establish that the axioms were not *categorical*, that is, that they did not uniquely characterize the objects of the intended mathematical domain (such as the integers or the real numbers).

Peano, on the other hand, had shown that his *second-order* axioms for the natural numbers *were* categorical. Accordingly, if some number-theoretic statement were shown to be undecidable from *those* axioms, it would mean that second-order logic was *not* complete, so that a set of axioms, even though noncontradictory, might have *no* realization whatever; and that would conclusively demonstrate the inadequacy of consistency as a criterion for mathematical existence.

4 FORMALLY UNDECIDABLE STATEMENTS, THE UNDEFINABILITY OF TRUTH, AND THE EXISTENCE OF NONSTANDARD MODELS

In the definitions of validity and satisfiability given in Hilbert and Ackermann's book, what it meant for an expression to be 'correct' was left undefined. Remarkably, that central semantic notion was also left undefined in Gödel's paper on completeness. In both texts it was simply assumed that there was a clear intuitive understanding of whether an interpreted formula held or not when particular objects were substituted for its free variables.

It was not long, however, before Gödel confronted the question whether the notion of truth, for an arbitrary formula of the language of first-order Peano arithmetic, could be formally defined *within* the language of arithmetic itself. The occasion for doing so, he later explained, was his attempt "to give a relative model-theoretic consistency proof of analysis in arithmetic", in the spirit of Hilbert's reductive program for establishing the consistency of higher-level mathematical theories relative to that of simpler ones.[3] Gödel quickly realized that though a formal definition of *provability* from Peano's axioms could be given within the language of first-order Peano arithmetic, that of *truth* within the domain of the natural numbers could not.[4]

Gödel exploited that difference to show that if Peano's axioms were consistent, one could exhibit a statement true of the natural numbers that was not formally derivable from the axioms. In view of his completeness theorem, any such statement would have to be false in some other structure that satisfied the first-order Peano axioms. So the first-order version of Peano's axioms, if consistent, could not be categorical.

Such was the genesis of Gödel's most famous result, his (first) *incompleteness theorem*, which he announced at the same time as his completeness result, during a conference at Königsberg in September of 1930. (See [Dawson, 1984] for further details of that momentous occasion.) The result appeared in print early the next year [Gödel, 1931], and was later accepted as Gödel's *Habilitationsschrift.* In that paper Gödel began by giving an informal précis of the main idea of his proof "without any claim to complete precision". The principal technical tool he employed was that of *arithmetization of syntax*, whereby all of the symbols of the formal language of number theory, as well as all sequences thereof and all sequences of *them*, were encoded in a systematic way as natural numbers. The class of well-formed formulas (sequences of symbols of a particular kind), of proofs

---

[3] Gödel to Yossef Balas, n.d., published on pp. 9–11 of [Gödel, 2003a].

[4] Evidence of his early awareness of those facts is provided by his letter to Balas, cited above, and by his letters of 2 April 1931 to Paul Bernays and of 12 October 1931 to Ernst Zermelo. The latter are reproduced on pp. 91–99 of [Gödel, 2003a] and pp. 423–429 of [Gödel, 2003b].

(sequences of formulas of a particular kind), of unary formulas (those with a single free variable), and of numerical substitution instances of the latter (in which the free variable is replaced by a numeral denoting a particular natural number) then correspond to certain sets of natural numbers, each of which can be shown to be definable[5] within the framework of formal number theory.

Because it is possible to order all the unary formulas in a definable sequence, within the theory a binary number-theoretic predicate $Q(x, y)$ can be defined such that if $F(x)$ is the nth unary formula and $\mathbf{n}$ is the numeral corresponding to n (that is, the term of the formal language obtained by applying the successor function n times to the constant symbol $\mathbf{0}$), the formula $Q(x, \mathbf{n})$ encodes the statement "x is not the code number of a proof of the formula $F(\mathbf{n})$." Quantifying universally over the variable x yields the unary formula $\forall x Q(x, y)$, which has some index number $\mathbf{q}$ in the ordering of all unary formulas. The formula $\forall x Q(x,\ \mathbf{q})$, henceforth abbreviated as S, then encodes the statement "the $\mathbf{q}$th numerical substitution instance of the qth unary formula is not provable". That is, S affirms its own unprovability. Gödel then argued informally that if the underlying theory were *consistent*, the sentence S could not be provable, and if it were *sound* (that is, if no false statement were provable — a stronger condition than consistency), neither could its negation.

The notion of consistency is syntactic (for no formula F are both F and its negation provable), but that of soundness involves the notion of truth, against which, Gödel believed, there was at that time "a prevailing prejudice".[6] Accordingly, in the precise analysis that he gave in the main body of his paper, he replaced the assumption that the theory was sound by that of its being *ω-consistent*, the purely syntactic criterion that if each numerical substitution instance $F(\mathbf{n})$ of a unary formula $F(y)$ is refutable, the existential statement $\exists y F(y)$ is not provable. The resulting (first) incompleteness theorem then stated: In any ω-consistent axiomatization of number theory that is strong enough to permit the arithmetization described above, if the set of formulas is primitively recursively definable, the sentence S constructed as above is neither provable nor refutable.

Any ω-consistent system is consistent, since, by definition, there is a formula that is not provable within it. But any system Σ that satisfies the hypotheses of the incompleteness theorem (and so is ω-consistent) can be extended by adjoining the negation of the undecidable statement S. The resulting system Σ^* must be consistent, since otherwise S would have been provable in Σ. However, S is the universal generalization of the formula $Q(\mathrm{x},\mathbf{q})$, whose nth numerical substitution instance $Q(\mathbf{n},\mathbf{q})$ asserts that the numeral corresponding to the natural number n is not the numeral for the code number of a proof in Σ of S. But in the course of his proof Gödel showed that $Q(\mathbf{n},\mathbf{q})$ is provable in Σ if n is not, in fact, the code number of a proof in Σ of S, and is *dis*provable in Σ if n *is* the code number of a proof in Σ of S. Thus if, for some n, the *negation* of $Q(\mathbf{n},\mathbf{q})$ were true, S would be refutable in Σ; hence no numeral **n** can correspond to the code number of a proof of S in Σ. Therefore $Q(\mathbf{n},\mathbf{q})$ must be provable in Σ, and hence in $\Sigma*$, for *every* n. But that means Σ^* is not ω-consistent, since the negation of S is also a theorem of Σ^* .

---

[5] More precisely, (primitively) recursively definable, a notion Gödel rigorously defined later in the paper.

[6] See the letter to Y. Balas cited above. Cf. also Rudolf Carnap's recollections in his *Intellectual Autobiography* [Carnap, 1963] and the *caveat* in section 9 of [Church, 1956] (quoted on p. 133 of [Henkin, 1996].

The incompleteness theorem holds for second-order formalizations of number theory as well as for first-order axiomatizations, so the state of affairs described at the end of the preceding section is actually the case: If Σ is taken to be second-order Peano arithmetic, then Σ^* is a consistent system that has no realization. On the other hand, if Σ is taken to be the first-order Peano axioms, then Σ^* is a consistent first-order theory. As such (by the completeness theorem) it must have a model, in which there must be an object that instantiates the negation of S. That object must satisfy the statement within the theory that asserts that it is the code number of a proof of S. But it cannot correspond to any *numeral* in the theory. Thus the model must contain *nonstandard* numbers.

5 THE SECOND INCOMPLETENESS THEOREM, AND LATER EXTENSIONS OF GÖDEL'S RESULTS

Following his announcement at Königsberg Gödel was drawn aside by John von Neumann, who was eager to learn further details about the existence of undecidable statements. Not long afterward, as we know from their correspondence [Gödel 2003b, 336–341], both men independently discovered the *second* incompleteness theorem, according to which, in any consistent, recursively axiomatizable extension $T$of Peano arithmetic, the natural formalization Con_T of the statement that T is consistent (a formula expressing that no code number can be that of a formula that is both provable and refutable) is itself formally unprovable within T. Thus, contrary to Hilbert's hope, the consistency of arithmetic cannot be demonstrated through reduction to that of a syntactically weaker system.[7]

Gödel announced the second theorem in an abstract [Gödel, 1930b] that he communicated to the Vienna Academy of Sciences on 23 October 1930, and he sketched its proof in the final section of his paper *1931*, which was received for publication on 20 November 1930. (Von Neumann wrote to apprise Gödel of his proof of the result three days later.) In the last sentence of that paper Gödel promised to give full details of the proof "in a sequel to be published soon". But no sequel ever appeared. A detailed proof was first published in 1939, in the second volume of Hilbert and Bernays' text *Grundlagen der Mathematik* [Hilbert and Bernays, 1939].

During the interim between the publication of the proofs of the first and second theorems a number of important extensions of Gödel's methods and results appeared. One was Tarski's proof, in a postscript added in press to his Polish monograph [Tarski, 1933], that the notion of truth for formulas of arithmetic is not definable within arithmetic itself. As noted above, Gödel was well aware of that fact, but he had refrained from stating it in print because he wished to avoid the controversy that he thought discussion of semantic issues would engender. Tarski was bolder, and in the monograph just cited (better known in its German translation [Tarski, 1935] he analyzed the concept of truth as it was used both in natural and formal languages and gave an inductive definition (within a second-order metalanguage) of the notion of truth in a formal language L. Tarski explicitly credited Gödel with having introduced the methods needed to establish the undefinability of truth, but he also noted that, independently of and prior to Gödel, he had given an example of a consistent theory that was not ω-consistent.

---

[7] One may, however, demonstrate the consistency of arithmetic relative to systems that appear more perspicuously unobjectionable, as Gerhard Gentzen and, later, Gödel himself, did. See §6 below.

In a later paper [Tarski, 1936], published the same year in German translation), Tarski analyzed the notion of logical consequence; and on the basis either of that paper or the 1933 monograph he has generally been credited with having given the modern model-theoretic definitions of logical truth and logical consequence. But in a detailed essay on Tarski's contributions to the study of those notions [Etchemendy, 1988] John Etchmendy has pointed out that neither of those attributions is correct. He notes in particular (pp. 65 and 69) that "as long as the quantifiers are treated as logical constants", the "semantic account of the logical notions" that Tarski gave in his 1936 paper "leaves the domain of quantification fixed", whereas the definition that has since become standard "requires that... the domain of quantification" be varied. Etchemendy concludes that the history of the model-theoretic analysis of logical truth and logical consequence "is probably too complex to support [its] attribution to any single author."

It was also in 1936 that Rosser strengthened the first incompleteness theorem by demonstrating that the hypothesis of ω-consistency is not needed in order to establish the existence of undecidable statements in formalized arithmetic [Rosser, 1936]. He did so by replacing Gödel's sentence that asserts its own unprovability by a more complex self-referential statement that asserts that *if* it is provable within the system, there is a proof of its negation with a smaller code number. Simple consistency of the underlying system then suffices to show that Rosser's statement is formally undecidable within it.

A third result related to Gödel's completeness and incompleteness theorems was obtained by Skolem in a paper he published in 1933 [Skolem, 1933]. By an explicit construction that foreshadowed the later notion of ultraproduct, Skolem demonstrated that there is a structure not isomorphic to the natural numbers that nevertheless satisfies exactly the same sentences of formal number theory that the natural numbers do. (Following Tarski, structures that satisfy exactly the same sentences are said to be *elementarily equivalent.*) Skolem drew as a corollary that no finite axiomatization of number theory can fully characterize the natural numbers, a result that, as noted above (and as Gödel observed in his review of Skolem's paper, [Gödel, 1934], in *Zentralblatt für Mathematik*), follows directly from the completeness and incompleteness theorems. Yet neither Gödel nor Skolem seems to have noticed that the main result of Skolem's paper is an easy consequence of another theorem that Gödel had proved in his 1930 paper: the (countable) *compactness theorem*, mentioned earlier, which asserts that if every finite subset of a (countably infinite) set of first-order statements is satisfiable, there is a structure that satisfies all the statements in the set.

Today the compactness theorem is considered one of the cornerstones of the theory of models. But that branch of logic did not become an organized subdiscipline until the 1950s, and it took many years before the significance of the compactness property came to be recognized. (Indeed, the very name "compactness", with its topological connotations, is due to Tarski and dates from 1952.) A detailed account of the rather convoluted history of the compactness theorem, in which the principal players, in addition to Gödel, were Henkin, Tarski, Abraham Robinson and, especially, A.I. Malt'sev, is given in [Dawson, 1993]. For Gödel himself the theorem seems to have been an afterthought. There is no mention of it in his dissertation, but only in the published revision thereof, where it is used as a lemma in the proof of the completeness theorem (in contrast to most modern treatments, where compactness is deduced as a *corollary* of completeness).

The easy non-constructive argument, invoking the compactness theorem, that establishes Skolem's main theorem proceeds by considering the set of all sentences that are true in the standard model, together with the statements $c > \mathbf{n}$, for each natural number n , where c is a single new constant symbol adjoined to the language of Peano arithmetic. It is this last move, that of expanding the language by adjoining a new constant symbol, that seems to have escaped notice for more than a decade after Gödel stated and proved the compactness theorem; and it was an extension of that method, involving the adjunction of infinitely many new constant symbols, that Henkin employed in his dissertation to give a new proof of the completeness theorem — one which, unlike Gödel's, could be extended to uncountable languages.

Henkin himself has given a detailed and fascinating account [Henkin, 1996] of how his completeness proof came about, as an unexpected offshoot of a failed attempt to establish a very different result. Given a consistent set Σ of first-order sentences, his key idea was to use added constant symbols to build a model for Σ. In particular, he saw how constant symbols could be chosen to serve as *witnesses* for the existential sentences of the expanded language, by enumerating all such sentences and associating to each sentence $\exists x A_n(x)$ a sentence T_n of the form $\exists x A_n(x) \rightarrow A_n(c)$, where c is a suitably chosen one of the new constants that does not appear in $A_n(x)$ nor in any of the sentences T_m previously introduced. If we denote by T the set of all the sentences T_n , then the set $\Sigma \cup T$ remains consistent, and, by a well-known theorem of Adolf Lindenbaum (first published in [Tarski, 1930], it may be extended to a maximal consistent set M of sentences.

The set of all closed terms in the enlarged language may then be taken as the domain of a structure A for interpreting that language. In that structure , each n-ary function symbol f of the language is interpreted by the n-ary function whose value at an n-tuple $(\tau_1, \tau_2, \ldots, \tau_n)$ of closed terms is the closed term $f\tau_1\tau_2\ldots\tau_n$.

Each n-ary relation symbol R of the language (including the equality symbol, if present) is interpreted in the structure by an n-ary relation R which holds of closed terms $\tau_1, \tau_2, \ldots, \tau_n$ if and only if $R\tau_1\tau_2 \ldots \tau_n$ is in M. And each constant symbol c is interpreted by itself. If the language does not contain the equality symbol, then A is a model for M. Otherwise, the quotient structure $\mathcal{A}/E$, where E is the relation that interprets the equality symbol, will be a model for M. In either case, the reduct of $\mathcal{A}$ or $\mathcal{A}/E$ to the original language is a model for Σ.

6 CONSISTENCY PROOFS IN THE WAKE OF THE SECOND INCOMPLETENESS THEOREM

Hilbert's aim in proposing his formalist program was to secure the foundations of mathematics by proving, by "finitary" means, the consistency of successively more complex axiom systems, beginning with Peano arithmetic. The challenge of carrying out the initial step in that program was the second of the problems that he posed in his famous address to the International Congress of Mathematicians in 1900. Gödel's second incompleteness theorem implied that if all "finitary means" (methods of inference that Hilbert never precisely delimited) can be formalized within Peano arithmetic, then that goal cannot be realized. But in his incompleteness paper Gödel explicitly stressed that it was "conceivable that there exist finitary proofs that *cannot* be expressed" in Peano arithmetic. As time went on, that possibility came to seem less and less likely. But the question remained: Can the consistency of arithmetic be established by means which, though not formalizable within arithmetic itself, nonetheless appear to be "more evident [or] reliable,... so that one's conviction is thereby strengthened"?[8]

The first such proof, based on the principle of transfinite induction up to the ordinal ε_0 (the limit of the sequence $\{s_n\}$ of ordinals given by $s_0 = \omega, s_{n+1} = \omega_{s_n}$), was given by Gerhard Gentzen in 1936 [Gentzen, 1936] and was carried out within the framework of a calculus for "natural deduction" that Gentzen had developed two years earlier [Gentzen, 1935].

The basic entities within that calculus are expressions of the form $\Gamma_1, \Gamma_2, \ldots, \Gamma_n \rightarrow \Lambda$, called *sequents*, where $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ (the *antecedents*) and Λ (the *succedent*) are formulas of an underlying formal language. The intended meaning of such a sequent, according to Gentzen, is that Λ follows from the *assumptions* $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$. Sequents of the form $\rightarrow \Lambda$, in which there are no assumptions, are also admitted; they are regarded as axioms.

As the term 'natural deduction' implies, Gentzen's original aim was to formalize logical deductions in a manner that mimicked, as closely as possible, the patterns of informal mathematical reasoning. Such reasoning is not ordinarily carried out within an axiomatic framework, but rather involves the drawing of conclusions from given premises. Sequents are a way of formalizing the steps involved in such deductions. In particular, Gentzen defined a *derivation*, within the calculus of sequents, to be a finite list of sequents, each of which is either a *basic sequent* (an axiom, or else a sequent of the form $\Lambda \to \Lambda$), is the result of applying one of a specified set of *structural transformations* to an earlier sequent (such as rearranging the order of its assumptions, or eliminating duplicated assumptions from it), or is obtained from an earlier sequent by the application of one of a finite number of *rules of inference*. For arithmetic, the latter comprised the rule of mathematical induction together with the logical rules *reductio ad absurdam*, elimination of a double negation, and five pairs of introduction and elimination rules, one for each of the binary connectives ($\&, \vee$ and $\supset$) and each of the quantifiers ($\forall$ and $\exists$) that Gentzen employed.

---

[8] The words are quoted from a lecture Gödel gave in January of 1938 [Gödel, 1938a], in which he surveyed the nature of consistency proofs and outlined a method for establishing the consistency of arithmetic via the notion of "computable functional of finite type" (a result he did not publish until twenty years later [Gödel, 1958]).

To establish the consistency of arithmetic, Gentzen had to show that no sequent of the form $\to (A\&\neg A)$ could occur as the last sequent in any derivation. He did so in several stages. First, he indicated how to replace the symbols $\vee, \exists$, and $\supset$, wherever they occurred in a derivation, by logically equivalent expressions involving $\neg$, $\&$, and $\forall$; second, he described how to modify the sequence of rules of inference accordingly, so as to obtain a derivation, employing only the restricted set of connectives and quantifiers, of the final one of the resulting sequents from the initial one; third, he described a set of *reduction rules* for sequents in the restricted language, whereby those of particular forms can be converted, in finitely many steps, into others of a certain *reduced form*; and finally, he extended the notion of reduction to entire derivations and showed that any derivation could be converted, again via finitely many reduction steps, into one whose last sequent was in reduced form.

The reduced form of a sequent contains no free variables; the formulas in it are quantifier-free numerical statements whose truth or falsity is decidable through numerical computation; and as the name implies, no further reduction rules are applicable to it. Moreover, every sequent in reduced form is *true*, in the sense that either its succedent is true or both its succedent and at least one of its antecedents is false. To complete the proof Gentzen then showed that if there were any derivation whose last sequent had the form $\to (A\&\neg A)$, there would be another derivation whose final sequent was $\to 1 = 2$. But the latter is a false sequent to which none of the reduction rules is applicable, so its derivation could not be converted to one in reduced form.

The details of Gentzen's proof are complicated.[9] The crucial step is that of showing that the reduction rules must cease to be applicable to a derivation after finitely many reduction steps. To do so, Gentzen proceeded by induction on the complexity of the derivations, which he measured by correlating them with ordinals $< \varepsilon_0$. It was at that point that his proof transcended the methods that are formalizable within arithmetic.

---

[9] Two years later he gave an alternative proof [Gentzen, 1938] that is generally deemed to be more readily comprehensible.

7 THE EMERGENCE OF RECURSION THEORY

Recursive computational algorithms — procedures for calculating output values from given input values in a finite sequence of steps, using values computed at previous steps as inputs to later steps — have been employed by mathematicians since ancient times. But the question of characterizing precisely *which* functions can be computed by recursive algorithms — for short, those that are *effectively computable* — did not arise until the twentieth century.

A first step in that direction was taken by Richard Dedekind in 1888, in his book *Was sind und was sollen die Zahlen?* (What are numbers, and what should they be?) There, in Satz 126, he provided a rigorous proof that number-theoretic functions whose values are specified by inductive definitions are well defined, and he went on to give the inductive definitions for addition and multiplication of natural numbers that Peano adopted in his axiomatization of arithmetic, published the following year. Four years later the first volume of Frege's *Grundgesetzte der Arithmetik* appeared, in which, as noted earlier, Frege attempted to derive arithmetic from logic; but after the discovery of Russell's paradox showed Frege's system to be contradictory, attention shifted to the "crisis" in foundations.

The theory of types, as expounded by Russell in *Principia Mathematica* (1910–13), offered a way to avoid the paradoxes, but it was cumbersome, and the axiom of reducibility, which Russell had felt impelled to introduce in order to overcome certain difficulties, in effect compromised the whole type structure. A better way of founding arithmetic was called for, and in 1919, after studying *Principia,* Skolem set out to develop arithmetic on the basis of "the recursive mode of thought" (that is, mathematical induction), without the use of bound variables ranging over infinite domains. The result was his paper [Skolem, 1923a], in which he developed the theory now known as *primitive recursive arithmetic.* Skolem used inductive definitions and proofs to derive the basic properties of addition, subtraction, ordering ($<$), multiplication, divisibility, primes and prime factorizations. Only bounded quantifiers were employed, and even they were introduced as eliminable abbreviations for finite conjunctions or disjunctions.

Skolem's work showed how far arithmetic could be developed along finitary lines. Skolem also proved that the apparently more general principles of definition and proof by course-of-values recursion (whereby the value $f(n+1)$ of a function is specified, or the truth of a proposition $P(n+1)$ is proved, with reference to the values of f or the truth of P at all natural numbers $\leq n$, not just at n itself) do not, in fact, go beyond what is obtainable by ordinary inductive definitions or proofs. But Skolem did not work within a formal framework, and though he gave many examples of number-theoretic functions and relations that could be defined inductively, he did not provide any *schema* for characterizing the class of all functions or relations that could be so obtained.

It was Gödel, in his incompleteness paper, who first gave a schematic characterization for the class of functions that he called *rekursiv.* He began by declaring a number-theoretic function $\phi(x_1, x_2, \ldots, x_n)$ to be *rekursiv definiert* (recursively defined) in terms of number-theoretic functions $\psi(x_1, x_2, \ldots, x_{n-1})$ and $\chi(x_1, x_2, \ldots, x_{n+1})$ if

$$\phi(0, x_2, \ldots, x_n) = \psi(x_2, \ldots, x_n)$$

and

$$\phi(k+1, x_2, \ldots, x_n) = \chi(k, \phi(k, x_2, \ldots, x_n), x_2, \ldots, x_n).$$

A function was then defined to be *rekursiv* if it was the last in a finite sequence of functions, each of which was either a constant function, the successor function $x+1$, was obtained from a previous function in the sequence by substituting other previous functions in the sequence for some of its arguments, or was recursively defined in terms of two of the preceding functions. An n-place relation R was defined to be *rekursiv* if, for some *rekursiv* function ϕ, $R(x_1, x_2, \ldots, x_n)$ held if and only if $\phi(x_1, x_2, \ldots, x_n) = 0$. Like Skolem, Gödel showed that some forty-five specific number-theoretic functions and relations were *rekursiv* (but that the relation Bew(x), which held of x if and only if x was the code number of a formula provable in formal number theory, was not). He also proved that every *rekursiv* relation was arithmetically definable, that is, could be defined over the domain of natural numbers by a first-order formula in the language with equality having functions symbols only for addition and multiplication.

The class of *rekursiv* functions was all that Gödel needed to establish his incompleteness results. In his 1931 paper he did not aim to characterize all number-theoretic functions that were effectively computable, and it was clear by then that the *rekursiv* functions did not comprise all that were. For in 1928 Wilhelm Ackermann had given an example of a computable function ξ of three natural-number variables, n, b, and a, defined by simultaneous recursion on n and b, for which it could be proved that $\xi(b, b, b) > \phi(b)$ for any unary *rekursiv* function. Specifically, starting from the function $\alpha(n, a)$ that equals 0 if $n = 0, 1$ if $n = 1$, and a if $n > 1$, Ackermann defined

$$\xi(0, b, a) = a + b,$$
$$\xi(n+1, 0, a) = \alpha(n, a),$$

and

$$\xi(n+1, b+1, a) = \xi(n, \xi(n+1, b, a), a).$$

It follows that $\xi(1, b, a) = a \cdot b, \xi(2, b, a) = a^b$, etc. Subsequently, in a series of papers, Rosza Péter studied the hierarchy of functions defined by n-fold simultaneous recursions, and in 1934 she introduced the term 'primitive recursive' for the class of functions that Gödel had called *rekursiv.*

In the meantime, in his papers [Church, 1932] and [Church, 1933], Church had proposed an alternative system of logic to which he hoped Gödel's incompleteness theorem would not apply. In the end that hope was dashed, for Kleene and Rosser, two of Church's own students, proved that the system was, in fact, inconsistent. Nevertheless, a subsystem of it, the *λ-calculus*, proved not only to be consistent, but to be unexpectedly powerful as a means for defining number-theoretic functions.

When Church created the λ-calculus he did not foresee its use in characterizing the class of effectively computable functions. Rather, he introduced it simply to provide a way to distinguish the *function* of x, defined for each x as the value at x of some formula f, from its *values.* For the former he wrote $\lambda f x \cdot f(x)$, for the latter $f(x)$. Given that λ-*notation*, the corresponding calculus then provided rules for evaluating functions by substituting expressions for their arguments (an operation Church called "reduction"), for expanding expressions so as to obtain others having the same numerical value (the reverse of reduction), and for changing the bound variables within expressions. To do arithmetic, Church identified the natural numbers $1, 2, 3, \ldots$ with the terms $\lambda f x \cdot f(x)$, $\lambda f x \cdot f(f(x)), \lambda f x \cdot f(f(f(x))) \ldots$ formed by functional iteration, where f denotes a particular (fixed) function symbol.

In [Church, 1933] the development of arithmetic within the λ-calculus, in particular the proof of Peano's axioms, was left as an open problem — one that, in January of 1932, Kleene chose as the subject for his doctoral dissertation. By his own later account [Kleene, 1981, 56], three of the five Peano axioms were easy to prove, but "the third axiom (if the successors of two numbers are equal, so are the numbers) offered a challenge." The chief difficulty was that of finding a way to define the predecessor function in the λ-calculus (where 1 was defined to be its own predecessor, and n the predecessor of $n+1$). After much thought, Kleene saw how to do so "one day late in January or early in February 1932, while in a dentist's office" (!). He promptly communicated the result to Church, who confessed that "he had just about convinced himself" that the predecessor function was not λ-definable. That it turned out to be, after all, "excited [Church and Kleene's] interest in what functions... [were] actually λ-definable", and "the exploration of [that problem] became a major subproject" of Kleene's dissertation [ibid, 56–57].

Thereafter events moved swiftly. In June of 1932, having read [Church, 1932] not long after its appearance, Gödel wrote to Church to inquire how it was possible to prove "absolute existence propositions" in his system, "for example the axiom of infinity"; and in his reply the following month Church mentioned that "a student... in Princeton" had recently proved the Peano axioms within the λ-calculus, a result that he himself had "checked . . . in detail" and that established "the essential part of the axiom of infinity".[10] Meanwhile, as Kleene recalled in the same memoir cited above, "from February 1932 on Church and I knew that only effectively calculable functions can be λ-definable. We kept thinking of specific such functions, and of specific operations for proceeding from such functions to others. I kept establishing the functions to be λ-definable and the operations to preserve λ-definability." Kleene submitted his dissertation in September of 1933, before the discovery of the inconsistency in Church's full system (which caused him to have to rewrite the dissertation the following spring), and shortly thereafter Gödel arrived in Princeton for the first of three visits that he made to the Institute for Advanced Study prior to his emigration there in 1940. Late that same year, as recalled in [Rosser, 1984], Church remarked to Rosser (though without "firm conviction") that "perhaps every effectively calculable function from positive integers to positive integers is [λ-definable]"; and by early in 1934 he had become convinced enough that that was the case to propose the idea to Gödel — the first version of what would later be called "Church's Thesis".

---

[10] The correspondence is reproduced on pp. 366–369 of [Gödel, 2003a].

Gödel, however, thought Church's proposal to be "thoroughly unsatisfactory" — after all, only two years before Church had thought something as elementary as the predecessor function was unlikely to be λ-definable — to which Church responded that if Gödel were to "propose any definition of effective calculability which seemed even partially satisfactory", he would endeavor to show that "it was included in λ-definability".[11] And that is exactly what happened.

From February to May of 1934 Gödel lectured at the I.A.S. on extensions of his incompleteness results. Notes on the lectures, taken by Kleene and Rosser, were distributed to subscribers and submitted at the end of the course to Gödel, who corrected and approved them.[12] The last section, entitled "General recursive functions", began with a variant of Ackermann's function , as an example of the "various sorts of functions" that could be "defined by inductions". Consideration of such functions, Gödel said, "leads to the question what one would mean by 'every recursive function"', and in the following paragraphs he drew upon a suggestion of Jacques Herbrand, communicated to him in a private letter[13], to suggest how one might "attempt to define" that notion: "If ϕ is an unknown function, $\psi_1, \ldots, \psi_k$ are known functions, and the ψ's and ϕ are substituted into one another in the most general fashions and certain pairs of the resulting expressions are equated" (subject to the restriction that the left-hand member of each equation have the form $\phi(\psi i_1(x_1, x_2, \ldots, x_n), \psi i_2(x_1, x_2, \ldots, x_n), \ldots, \psi i_k(x_1, x_2, \ldots, x_n))$), then ϕ is a general recursive function "if the resulting set of functional equations has one and only one solution for ϕ" (assuming in addition that "all possible set of arguments …of ϕ can be so arranged that the computation of the value of ϕ for any given set of arguments …requires a knowledge of the values of ϕ only for [preceding] sets of arguments").

As Gödel later took pains to explain to Martin Davis (see [Davis, 1982], at the time of his lectures he was still "not at all convinced" that the definition of general recursive function he had proposed would suffice to "comprise all possible recursions". And he remained unconvinced a year later, by which time Rosser and Kleene, independently, had proved that all functions that are general recursive according to Gödel's definition are λ-definable, and Church and Kleene, again independently, had demonstrated the converse. (See [Sieg, 1997] for a more detailed and precise chronology of those events.) For Church, though, the fact that the two proposed characterizations had proved to be coextensive was apparently what emboldened him to proclaim his thesis publicly. He did so in a lecture to the American Mathematical Society on 19 April 1935, whose abstract, containing a statement of the thesis, was received on 22 March. What is remarkable, and has been examined in detail in [Gandy, 1988; Sieg 1997; Soare 1996], is that in his talk and his subsequent publication, [Church, 1936a] Church did not propose that the effectively computable functions be identified with those that are λ-definable (which, as noted earlier, he had not introduced with that end in mind), but with *Gödel's* class of general recursive functions. It seems that Church, Kleene and Rosser all felt that "general recursivity... embod[ied] the idea of effective calculability" more perspicuously than did λ-definability [Rosser,1984, 345].

---

[11] The quotations are from Church's letter to Kleene of 29 November 1935, as reported in [Kleene, 1981, 59].

[12] They were belatedly published (with a Postscript dated 3 June 1964) in [Davis, 1965] (reprinted in [Gödel, 1990, 346–371].

[13] Found by the author in Gödel's *Nachlass* after his death and published in [Gödel, 2003b, 14–21]. Herbrand's letter to Gödel is dated 7 April 1931, less than four months before Herbrand's untimely death in a mountaineering accident.

Yet another concept coextensive with that of λ-definability, μ-recursiveness, also came out of Kleene's research. As Kleene recalled, "In racking my brains for still more examples of effective definitions on which to try out my ability to λ-define, I thought of the least-number operator 'the least y such that' ", which he later denoted by μy. He observed that "If $R(x,y)$ is any ... effectively decidable relation", defined for all x and y , "and for each x there is a y such that $R(x,y)$" then the function given by $\mu y R(x,y)$ is defined for all x and may be considered to be effectively calculable, even if no bound on y can be given in advance [Kleene, 1981, 58]. He proceeded to show that any such function was λ-definable (hence general recursive), and went on [Kleene, 1936] to prove his Normal Form Theorem, which established the converse: For any general recursive function $\phi(x)$ there is a ternary primitive recursive predicate $T(e,x,y)$, a unary primitive recursive function $U(x)$, and a numerical index e (a code number of a set of equations defining ϕ) such that for every $x, \phi(x) = U(\mu y T(e,x,y))$.

For modern logic as a whole, and recursion theory in particular, 1936 was an *annus mirabilis.* Not only did Gentzen's consistency proof for arithmetic, Rosser's improvement of Gödel's incompleteness theorem, and the paper by Kleene just cited all appear that year, but Church (in the same paper in which he proposed his 'Thesis') introduced the term *recursively enumerable set* for a set that is the range of a general recursive function, and, in a second paper [Church, 1936b], demonstrated the unsolvability of the decision problem for validity[14]; almost simultaneously, Turing published his landmark analysis [Turing, 1936] of the process of idealized human computation (culminating in his description of the abstract machines that now bear his name), together with his own independent proof of the unsolvability of the decision problem; and Post, in his [1936], described what he called 'finite combinatory processes', which, though they were conceived independently, bore a close resemblance to Turing's machines.

Turing's paper was a *tour de force* that swept away almost all of the remaining opposition to Church's Thesis. In it Turing broke down the process of human paper-and-pencil computation into elemental discrete steps whose effectiveness was beyond doubt. He then modeled the process of human computation by deterministic finite-state machines capable of reading the symbols of a finite alphabet written in squares on an unbounded paper tape, moving along the tape a square at a time in either direction, and writing symbols in blank squares on the tape. The behavior of the machines was specified by a finite set of instructions (program), which could itself be encoded by symbols on a tape, and Turing showed how to construct a *universal* machine that, given the encoded program for *any* one of his machines, would simulate that machine's behavior. He went on to show that no machine could determine which statements of the restricted functional calculus were formally derivable, thereby demonstrating (as Church had in terms of general recursive functions) that the decision problem for validity was unsolvable. And in an appendix that he added upon learning of Church's work, he proved that any computable function (one whose outputs are produced by one of his machines, given the corresponding input) is λ-definable, and conversely. (In contrast, as pointed out in [Soare, 1996, 300], "Post did not attempt to prove that his formalism coincided with any other such as general recursiveness", though he "expressed the expectation that... would turn out to be true", nor did he give any "analysis... of why the intuititively computable functions are computable in his system"; and there is nothing in Post's paper that corresponds to Turing's universal machine.)

---

[14] Actually, for derivability in the restricted functional calculus. The result for validity then follows from Gödel's completeness theorem.

The 'confluence of ideas' — the phrase Gandy [1988] used to describe the provable coextensionality of the separately conceived notions of λ-definability, general recursiveness, μ-recursivenes and Turing computability — certainly suggested that *some* natural class of functions had been formally characterized by those notions. But it was Turing's detailed analysis of the steps involved in human computation that ultimately convinced Gödel and compelled general agreement that the functions in question were, in fact, those that were effectively computable.

Nevertheless, as noted in [Davis, 1982], so long as the only computable functions under consideration were *total* functions (that is, had all the natural numbers as their domain), constructivists could still object to circularity in any of the equivalent definitions. Those objections were removed, however, by Kleene's introduction (in his [1938]) of *partial* recursive functions, which might be undefined for some natural numbers (corresponding to Turing machines that, for some inputs, failed ever to halt and produce an output). Instead of being circular, the notion of partial recursive function made manifest that the problem of deciding *which* systems of equations define total functions (or equivalently, which Turing machines halt for all inputs) is algorithmically unsolvable. In the same paper Kleene also presented his Recursion (or Fixed Point) Theorem, which, following [Soare, 1996], may be stated as follows: If ϕ_n is the partial recursive function computed by the nth Turing machine program, P_n, in some effective listing of all such programs, and if $f(x)$ is any *total* recursive function, there is a fixed point k for which $\phi_{f(k)=}\phi_k$.

By 1940 the principal concepts and theorems of 'ordinary' computability theory had been enunciated, and Turing, in his [1939], had planted the seed for their extension to 'relative' computability, through his introduction of the concept of an *oracle* (a source that a Turing machine might consult in order to obtain finitely many values of a non-computable function). Given the notion of an oracle, a set A is said to be recursive *in* B (or *relative to* B) if membership in A is computable from an oracle for membership in the set B. The relation "A is recursive in B" then yields a partial ordering $A \leq_T B$ (introduced in [Post, 1948]) whose equivalence classes, referred to as *degrees of unsolvability*, became the subject of much subsequent research.

8 THE RELATIVE CONSISTENCY OF THE AXIOM OF CHOICE AND THE GENERALIZED CONTINUUM HYPOTHESIS

Although the definition of general recursiveness that Gödel gave in his 1934 lectures was a stimulus to the development of recursion theory, Gödel himself was a bystander to that development — in part, it seems, because he did not foresee that it would be possible to give a definitive characterization of the intuitive notion of effective computability. Indeed, in his remarks before the Princeton Bicenntennial Conference on Problems in Mathematics, where he publicly acceded to Church's Thesis, he expressed his amazement that it had proved possible to give "an absolute definition" (one independent of the formalism chosen) for such an "interesting epistemological notion", deeming it "a kind of miracle... [that] the diagonal procedure d[id] not lead outside the defined notion" [Gödel, 1946, 1]. But beyond that, even before his visit to America in 1933–34, Gödel's interests had shifted to set theory. His incompleteness results had settled Hilbert's second problem, but the first problem remained open: To determine whether every infinite set of natural numbers is in one-to-one correspondence either with the natural numbers or with the continuum of real numbers, and whether that continuum can be well-ordered.

A number of relevant techniques and partial results bearing on those questions had already been obtained. Cantor had shown that the cardinality of any set was strictly less than that of its power set, and it was readily seen that the real numbers, as limits of countable sequences of rationals, could be put in one-to-one correspondence with the power set of the natural numbers. So the *continuum hypothesis* could be recast as the assertion that no set of natural numbers has cardinality strictly between that of the natural numbers and the reals. (The assertion that for *any* infinite set A, no set has cardinality strictly between that of A and its power set is the *generalized continuum hypothesis* [GCH], first proposed by Felix Hausdorff in 1930.) In 1904, and again in 1908, Zermelo had proved that if the axiom of choice were assumed, *any* set could be well-ordered [Zermelo, 1904; Zermelo, 1908a]; and in the latter year he had also put forward his (informal) axiomatization of set theory [Zermelo, 1908b]. The axiom of choice had then become the focus of criticism, together with the vague notion of *definit* property that Zermelo had employed in his axiom of separation.

In 1922 A.A. Fraenkel and Skolem, independently, realized that Zermelo's axioms did not suffice to allow consideration of the function F whose domain and initial value $F(0)$ were the set of natural numbers and whose value $F(n)$, for $n > 0$, was the nth iterate of the power set of the natural numbers. In particular, the range of F could not be shown to be a set on the basis of Zermelo's axioms, a defect that [Fraenkel, 1922] proposed to remedy by introducing the *axiom of replacement*.

In the more precise formulation given by Skolem in his [1923b], Zermelo's notion of *definit* property was interpreted as a relation that was *first-order definable* within the language of set theory; both the axiom of replacement and that of separation can then be expressed as infinite axiom *schemas*, each instance of which refers to a particular first-order formula ϕ: For a unary formula $\phi(x)$, separation asserts that for any set a the elements x of a for which $\phi(x)$ hold also constitute a set, while for a binary formula $\phi(x, y)$, replacement asserts that if $\phi(x, y)$ is a univocal relation, the image under ϕ of any set must itself be a set.

In 1925 von Neumann proposed an alternative version of set theory that interpreted *definit* properties in terms of *proper classes* — collections too large to be treated as sets, but which could nonetheless be handled without antinomy so long as they were not allowed to be *elements* of other classes; he achieved thereby a finite first-order axiomatization.[15] In the same paper [von Neumann, 1925] he also introduced the notion of an *inner model* of set theory — one whose domain consists of a class of sets that is first-order definable within a given model of the theory — and in [von Neumann, 1929] he used such a model to demonstrate that if his axioms were consistent, they would remain so if augmented by an axiom that prohibited the existence of infinite descending $\in$-chains of sets.

Such chains were allowed in Zermelo's 1908 axiomatization, as were so-called *urelements*: objects that were not themselves sets, but that could be elements of sets. The desirability of allowing urelements and/or infinite descending $\in$-chains was (and to some extent remains) a matter of debate. As usually formulated today, axiomatic set theory excludes both. But in 1922 Fraenkel showed that if urelements are permitted, they can be used to demonstrate the *independence* of the axiom of choice from the other axioms of set theory [Fraenkel, 1922]; and recently non-wellfounded sets, introduced by Peter Aczel in his *1988*, have found a number of applications in logic and computer science. See, e.g., [Barwise, 1996].

In 1930 Zermelo presented a revised axiomatization of set theory [Zermelo, 1930] that incorporated a (second-order) version of the replacment axiom as well as a new axiom, that of *foundation*, that disallowed infinite descending $\in$-chains. As in the first version, urelements were permitted, and starting from them Zermelo showed that by iterating the power-set operation transfinitely, all sets could be generated. (Alternatively, if urelements are disallowed, one may begin with the null set and define the α^{th} level of the *rank hierarchy* , $R(\alpha)$, by transfinite induction, defining $R(\alpha + 1)$ to be the power set of $R(\alpha)$, and $R(\beta)$, for limit ordinals β, to be the union of all $R(\alpha)$ of rank $\alpha < \beta$.) The first-order formulation of Zermelo's revised system (with schemas for separation and replacement), excluding urelements and the axiom of choice, subsequently became the standard axiomatization, *Zermelo-Fraenkel set theory* (ZF).

It appears that Gödel had begun thinking about Hilbert's first problem as early as 1928. In 1931 he studied Fraenkel's paper *1922* in detail, and in the fall of 1935, when he returned briefly to the I.A.S., he told von Neumann that he had succeeded in demonstrating the consistency of the axiom of choice relative to the other axioms of set theory. But almost immediately thereafter he was incapacitated by a serious bout of depression, from which he did not fully recover until the summer of 1937; and it was only then[16] that he succeeded in demonstrating the relative consistency of the generalized continuum hypothesis as well.

---

[15] In particular, the existence of classes $\{x : \phi(x)\}$ representing the extensions of first-order formulas is achieved using only a finite number of *comprehension axioms*.

The two consistency results were published in a series of papers. An announcement of both, in the context of von Neumann's system, was first made in [Gödel, 1938b], together with the remark that corresponding theorems held also for Fraenkel's system. The note [Gödel, 1939a] contained a very brief outline of how a model for the GCH could be formally constructed within any of "the system[s] of set theory (as formulated by A. Fraenkel, J. von Neumann, T. Skolem, [or] P. Bernays)". [Gödel, 1939b] gave "a sketch of the consistency proof[s] ... for Zermelo's [1908] axioms of set theory", and the monograph [Gödel, 1940] provided a self-contained development of set theory, including the consistency results, in a formally more streamlined but intuitively less perspicuous framework based essentially on the system in [Bernays, 1937].

The key idea, no matter which of the axiom systems is employed, is that of modifying the definition of the cumulative hierarchy so as to obtain another hierarchy. Specifically, define $L(0)$ to be the null set, and let $L(\alpha+1)$ be the set of all subsets of $L(\alpha)$ that are definable by first-order formulas $\phi(x)$, possibly containing constants for elements of $L(\alpha)$. At limit ordinals β, define $L(\beta)$ to be the union of the $L(\alpha)$ for $\alpha < \beta$. The union of all levels of the *constructible hierarchy* so defined forms a definable class L (a proper class in the systems of von Neumann or Bernays, or the extension of a predicate $C(x)$ in those of Zermelo, Fraenkel or Skolem), and since formulas and finite sequences of ordinals can be enumerated, the constructible sets possess a definable well-ordering. Crucially, the definition of the constructible hierarchy is also *absolute*, in the sense that given any model M of the underlying set theory, the sets constructible in M form a submodel L_M of M, and the sets constructible within that submodel are the same as those of L_M itself. That is, constructibility within the submodel is the same as constructibility in the original model. The elements of L_M are thus definably well-ordered, whence L_M satisfies the axiom of choice.

The proof that the GCH also holds within the constructible submodel is more difficult. If the *order* of a constructible set x is defined to be the least ordinal α for which $x \in L(\alpha+1)$, then Cantor's theorem on the cardinality of power sets implies that for each ordinal γ between ω_α and $\omega_{\alpha+1}$ there are constructible subsets of $L(\omega_\alpha)$ of order γ or greater. Gödel was able to show, however, that there are no subsets of $L(\omega_\alpha)$ of order $\omega_{\alpha+1}$. In L_M, therefore, the power set of ω_α, which is contained within the power set of $L(\omega_\alpha)$, must have cardinality $\leq \aleph_{\alpha+1}$.

In the early 1940s Gödel also sought to demonstrate the independence of the axiom of choice and the GCH in set theory without urelements. But it would take another two decades before Paul Cohen finally succeeded in that quest.

---

[16] Specifically, during the night of 14/15 June, according to a note in one of Gödel's workbooks.

9 CULMINATION AND CODIFICATION

By 1950 the major branches of modern logic, with the exception of model theory, had largely assumed their modern form. First-order logic had become the focus of study. The questions that Hilbert had posed to logicians in his 1900 address and his 1928 text with Ackermann had for the most part been answered (though not, in many cases, in the way he had expected). The future direction of proof-theoretic studies had been established through the work of Gödel and Gentzen. Recursion theory had become an important and highly technical specialty, poised to play a central role in the nascent field of computer science. The foundational controversies had died down. And set theory had become widely regarded as the basis for all of mathematics.

Model theory had yet to coalesce into an organized discipline, but many of the key results and techniques — the downward and upward Löwenheim-Skolem theorems, the compactness theorem, the method of elimination of quantifiers, ultraproducts, and the method of constants — had all been developed or foreshadowed, and Malt'sev, Henkin, Tarski and Robinson had begun applying model-theoretic techniques to the solution of questions in algebra.

At mid-century the time had come for modern logic to be codified in textbooks. With the appearance, within the next few years, of Kleene's *Introduction to Metamathematics* [1952] and Church's *Introduction to Mathematical Logic* [1956], that goal was achieved and modern logic may be said to have come of age.

BIBLIOGRAPHY

[Aczel, 1988] P. Aczel. *Non-well-founded Sets.* (Stanford, CA: Center for the Study of Language and Information. CSLI Lecture Notes No. 14), 1988.

[Barwise and Moss, 1996] J. Barwise and L. Moss. *Vicious Circles: On the Mathematics of Non-wellfounded Phenomena.* (Stanford, CA: Center for the Study of Language and Information), 1996.

[Bernays, 1937] P. Bernays. A system of axiomatic set theory. *The Journal of Symbolic Logic* 2: 65–77, 1937.

[Carnap, 1963] R. Carnap. Intellectual autobiography. In *The Philosophy of Rudolf Carnap*, ed. Paul Arthur Schilpp (La Salle, Illinois: Open Court Publishing Co.), 3–84, 1963.

[Church, 1932] A. Church. A set of postulates for the foundation of logic. *Annals of Mathematics (2)* 33: 346– 366, 1932.

[Church, 1933] A. Church. A set of postulates for the foundation of logic (second paper). *Annals of Mathematics (2)* 34: 839–864, 1933.

[Church, 1936a] A. Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics* 58: 345–363, 1936. Reprinted in [Davis, 1965, 88–107].

[Church, 1936b] A. Church. A note on the Entscheidungsproblem. *The Journal of Symbolic Logic* 1: 40–41 and 101–102 (*corrigenda*), 1936. Reprinted in [Davis 1965, 108–114].

[Church, 1956] A. Church. *Introduction to Mathematical Logic, I.* (Princeton: Princeton University Press), 1956.

[Davis, 1965] M. Davis. *The Undecidable.* (Hewlett, New York: Raven Press), 1965.

[Davis, 1982] M. Davis. Why Gödel didn't have Church's thesis. *Information and Control* 54: 3–24, 1982.

[Dawson, 1984] J. W. Dawson, Jr. Discussion on the foundation of mathematics. *History and Philosophy of Logic* 5: 111–129, 1984.

[Dawson, 1993] J. W. Dawson, Jr. The compactness of first-order logic: From Gödel to Lindström. *History and Philosophy of Logic* 14: 15–37, 1993.

[Etchemendy, 1988] J. Etchemendy. Tarski on truth and logical consequence. *Journal of Symbolic Logic* 53(1): 53–79, 1988.

[Fraenkel, 1922] A. Fraenkel. Der Begriff 'definit' und die Unabhängigkeit des Auswahlaxioms, *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalischmathematische Klasse*, 253–257, 1922. English translation in [van Heijenoort 1967, 284–289].

[Frege, 1879] G. Frege. *Begriffschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* (Halle: Nebert). 1879. English translation in [van Heijenoort, 1967, 1–82].

[Frege, 1893] G. Frege. *Grundgesetzte der Arithmetik, begriffsschriftlich abgeleitet.* Vol. 1. (Jena: Pohle), 1893.

[Gandy, 1988] R. Gandy. The confluence of ideas in 1936. In *The universal Turing machine: a half-century survey*, ed. Rolf Herken, (Oxford: Oxford University Press), 55–111, 1988.

[Gentzen, 1935] G. Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift* 39: 176–210 and 405–431, 1935. English translation in [Gentzen 1969, 68–131].

[Gentzen, 1936] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen* 112: 493– 565, 1936. English translation in [Gentzen 1969, 132–213].

[Gentzen, 1938] G. Gentzen. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften* 4 (n.s.): 19–44, 1938. English translation in [Gentzen, 1969, 252–286].

[Gentzen, 1969] G. Gentzen. *The Collected Papers of Gerhard Gentzen.* ed. M.E. Szabo. (Amsterdam: North-Holland Publishing Co.), 1969.

[Gödel, 1929] K. Gödel. Über die Vollständigkeit des Logikkalküls. Doctoral dissertation, University of Vienna, 1929. Reprinted, with English translation, in [Gödel, 1986, 60–101].

[Gödel, 1930a] K. Gödel. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik* 37: 349–360, 1930. Reprinted, with English translation, in [Gödel, 1986, 102–123].

[Gödel, 1930b] K. Gödel. Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit. *Anzeiger der Akademie der Wissenschaften in Wien* 67: 214–215, 1930. Reprinted, with English translation, in [Gödel, 1986, 140–143].

[Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik* 38: 173–198, 1931. Reprinted, with English translation, in [Gödel, 1986, 144–195].

[Gödel, 1934] K. Gödel. Review of *Skolem 1933. Zentralblatt für Mathematik und ihre Grenzgebiete* 7: 193–194, 1934. Reprinted, with English translation, in [Gödel, 1986, 379–381].

[Gödel, 1938a] K. Gödel. Vortrag bei Zilsel, 1938. Original and English translation in [Gödel, 1995, 86– 113].

[Gödel, 1938b] K. Gödel. The consistency of the axiom of choice and of the generalized continuum hypothesis. *Proceedings of the National Academy of Sciences, U.S.A.* 24: 556–557, 1939. Reprinted in [Gödel, 1990, 26].

[Gödel, 1939a] K. Gödel. The consistency of the generalized continuum hypothesis. *Bulletin of the American Mathematical Society* 45: 93, 1939. Reprinted in [Gödel, 1990, 27].

[Gödel, 1939b] K. Gödel. Consistency proof for the generalized continuum hypothesis. *Proceedings of the National Academy of Sciences, U.S.A.* 25: 220–224, 1939. Reprinted in [Gödel, 1990, 28–32].

[Gödel, 1940] K. Gödel. *The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory.* Annals of Mathematics Studies, Vol. 3., ed. Robert C. Gunning, John C. Moore, and Marston Morse. (Princeton: Princeton University Press), 1940. Reprinted in [Gödel, 1990, 33–101].

[Gödel, 1946] K. Gödel. Remarks before the Princeton Bicentennial Conference on problems of mathematics, 1946. In [Davis, 1965, 84–88]. Reprinted in [Gödel, 1990, 150–153].

[Gödel, 1958] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica* 12: 280–287, 1958. Reprinted, with English translation, in [Gödel, 1990, 240–252].

[Gödel, 1986] K. Gödel. *Collected Works.* Vol. I: Publications, 1929–1936, ed. Solomon Feferman et alii. (New York and Oxford: Oxford University Press), 1986.

[Gödel, 1990] K. Gödel. *Collected Works.* Vol. II: Publications, 1938–1974, ed. Solomon Feferman et alii. (New York and Oxford: Oxford University Press), 1990.

[Gödel, 1995] K. Gödel. *Collected Works.* Vol. III: Unpublished Essays and Lectures, ed. Solomon Feferman et alii. (New York and Oxford: Oxford University Press), 1995.

[Gödel, 2003a] K. Gödel. *Collected Works.* Vol. IV: Correspondence A – G, ed. Solomon Feferman et alii. (Oxford: Oxford University Press), 2003.

[Gödel, 2003b] K. Gödel. *Collected Works.* Vol. V: Correspondence H – Z, ed. Solomon Feferman et alii. (Oxford: Oxford University Press), 2003.

[Goldfarb, 1979] W. D. Goldfarb. Logic in the twenties: The nature of the quantifier. *The Journal of Symbolic Logic* 44: 351–368, 1979.

[Henkin, 1996] L. Henkin. The discovery of my completeness proofs. *The Bulletin of Symbolic Logic* 2(2): 127–158, 1996.

[Hilert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik.* (Berlin: Julius Springer Verlag), 1928.

[Hilbert and Bernays, 1939] D. Hilbert and P.Bernays 1939 *Grundlagen der Mathematik.* Vol. II. (Berlin: Springer), 1939.

[Kleene, 1936] S. C. Kleene. General recursive functions of natural numbers. *Mathematische Annalen* 112: 727–742, 1936. Reprinted in [Davis, 1965, 236–253].

[Kleene, 1938] S. C. Kleene. On notation for ordinal numbers. *The Journal of Symbolic Logic* 3: 150–155, 1938.

[Kleene, 1981] S. C. Kleene. Origins of recursive function theory. *Annals of the History of Computing* 3: 52–67, 1981.

[Löwenheim, 1915] L. Löwenheim. 1915 Über Möglichkeiten im Relativkalkul. *Mathematische Annalen* 76: 447–470, 1915. English translation in [van Heijenoort, 1967, 228–251].

[Moore, 1988] G. H. Moore. The emergence of first-order logic. In *History and Philosophy of Modern Mathematics*, ed. William Aspray and Philip Kitcher (Minneapolis: University of Minnesota Press), 95–135, 1988.

[Peirce, 1885] C. S. Peirce. On the algebra of logic: A contribution to the philosophy of notation. *American Journal of Mathematics* 7: 180–202, 1885.

[Post, 1921] E. L. Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics* 43: 169–173, 1921.

[Post, 1936] E. L. Post. Finite combinatory processes – formulation I. *The Journal of Symbolic Logic* 1: 103–105, 1936. Reprinted in [Davis, 1965, 288–291].

[Post, 1948] E. L. Post. Degrees of recursive unsolvability: preliminary report. *Bulletin of the American Mathematical Society* 54: 641–642, 1948.

[Rosser, 1936] J. B. Rosser. Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic* 1: 87–91, 1936. Reprinted in [Davis, 1965, 230–235].

[Rosser, 1984] J. B. Rosser. Highlights of the history of the lambda-calculus. *Annals of the History of Computing* 6: 337–49, 1984.

[Sieg, 1997] W. Sieg. Step by recursive step: Church's analysis of effective calculability. *The Bulletin of Symbolic Logic* 3: 154–180, 1997.

[Skolem, 1923a] T. Skolem. Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich. *Skrifter utgit av Videnskapsselskapet i Kristiana, I. Matematisk-naturvidenskabelig klasse* 6: 1–38, 1923. English translation in [van Heijenoort, 1967, 302–333].

[Skolem, 1923b] T. Skolem. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Matematikerkongressen i Helsingfors 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, 217–232. (Helsinki: Akademiska Bokhandlen), 1923. English translation in [van Heijenoort, 1967, 290–301].

[Skolem, 1929] T. Skolem. Über einige Grundlagenfragen der Mathematik. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Matematisk-naturvidenskapelig klasse* 7: 1–49, 1929.

[Skolem, 1933] T. Skolem. Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems. *Norsk matematisk forenings skrifter* 2(10): 73–82, 1933.

[Soare, 1996] R. I. Soare. Computability and recursion. *The Bulletin of Symbolic Logic* 2: 284–321, 1996.

[Tarski, 1930] A. Tarski. Über einige fundamentale Begriffe der Metamathematik. *Sprawozdania z posiedzen Towarzystwa Naukowego Warszawskiego, wydzial III* 23: 22–29, 1930.

[Tarski, 1933] A. Tarski. Pojecie prawdy w jezykach nauk dedukcyjnych. *Prace Towarzystwa Naukowego Warszawskiego, wydial III*, no. 34, 1933.

[Tarski, 1935] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia philosophica (Lemberg)* 1: 261–405. 1935.

[Tarski, 1936] A. Tarski. O pojeciu wynikania logicznego. *Przeglad Filozoficzny* 39: 58–68, 1936.

[Turing, 1936] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society (2)* 42: 230–265, 1936. Reprinted in [Davis, 1965, 115–153].

[Turing, 1939] A. M. Turing. Systems of logic based on ordinals, part 3. *Proceedings of the London Mathematical Society* 45:161–228, 1939. Reprinted in [Davis, 1965, 154–222].

[van Heijenoort, 1967] J. van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931.* (Cambridge, MA: Harvard University Press), 1967.

[van Heijenoort and Dreben, 1986] J. van Heijenoort and B. Dreben. Introductory note to [Gödel 1929; 1930] and [1930a]. In [Gödel, 1986, 44–59].

[von Neumann, 1925] J. von Neumann. Eine Axiomatisierung der Mengenlehre. *Journal für die reine und angewandte Mathematik* 154: 219–240, 1925. English translation in [van Heijenoort, 1967, 393–413].

[von Neumann, 1929] J. von Neumann. Über eine Widerspruchfreiheitsfrage in der axiomatischen Mengenlehre. *Journal für die reine und angewandte Mathematik* 160: 227–241, 1929.

[Wang, 1974] H. Wang. *From Mathematics to Philosophy.* (London: Routledge and Kegan Paul), 1974.

[Whitehead and Russell, 1910] A. N. Whitehead and B. Russell. *Principia Mathematica.* (Cambridge, U.K.: Cambridge University Press), 1910.

[Zermelo, 1904] E. Zermelo. Beweis, daß jede Menge wohlgeordnet werden kann. *Mathematische Annalen* 59: 514–516, 1904. English translation in [van Heijenoort, 1967, 139–141].

[Zermelo, 1908a] E. Zermelo. Neuer Beweis für die Möglichkeit einer Wohlordnung. *Mathematische Annalen* 65: 107–128, 1908. English translation in [van Heijenoort, 1967, 183–198].

[Zermelo, 1908b] E. Zermelo. Untersuchungen über die Grundlagen der Mengenlehre. I. *Mathematische Annalen* 65: 261–281, 1908. English translation in [van Heijenoort 1967, 199–215].

[Zermelo, 1930] E. Zermelo. Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae* 16: 29–47, 1930.

[Zermelo, 1932] E. Zermelo. Über Stufen der Quantifikation und die Logik des Unendlichen. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 41, pt. 2: 85–88, 1932.

INFINITY

Peter Fletcher

1 INTRODUCTION

Infinity occurs in many shapes and forms in mathematics. The points at infinity in projective geometry are very different from the infinite and infinitesimal quantities that occur in nonstandard analysis, or the transfinite numbers in set theory, or the infinity involved in a limiting process $\lim_{n\to\infty} a_n$. To classify this variety, it is helpful to distinguish between *actual* infinity and *potential* infinity (the distinction originates with Aristotle, who however meant it in a narrower sense than we do today (Lear, 1979–80)). According to the idea of actual infinity, infinite and finite quantities are subsumed under the same theory; an actually infinite quantity is just like a finite quantity, only bigger. According to the idea of potential infinity, infinity is merely a figure of speech: whenever we speak of infinity we are really talking about arbitrarily large finite quantities.

For example, the concept of natural number can be understood in terms of a generating process $0 \mapsto 1 \mapsto 2 \mapsto 3 \mapsto \cdots$, which can be continued as long as one pleases without coming to an end; this is potential infinity. Or it can be understood in terms of the set of all natural numbers $\mathbf{N} = \{0, 1, 2, 3, \ldots\}$, thought of as given once and for all; this is actual infinity. At first sight it may appear as if these two views come to the same thing, but we shall see in §3 that the difference is fundamental and its consequences pervasive.

To take another example, consider the sum of an infinite series, $\sum_{n=1}^{\infty} a_n$. From an actual-infinity viewpoint this would be understood as an addition of infinitely many quantities (think for example of infinitely many blobs of water that are lumped together into a single big blob); the total may of course be finite or infinite. Such a conception has some appeal in the case of physical quantities, as we shall see in §4.4. Similarly, an integral $\int_a^b f(x)\,dx$ can be seen as the addition of infinitely many infinitesimal areas $f(x)\,dx$. From this perspective, the three equations $\sum_{n=1}^{N} a_n + b_n = \sum_{n=1}^{N} a_n + \sum_{n=1}^{N} b_n$, $\sum_{n=1}^{\infty} a_n + b_n = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$, and $\int_a^b f(x) + g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$ are instances of the same rule (as the notation suggests).

This is not, however, the way infinite series are conventionally understood in modern mathematics. $\sum_{n=1}^{\infty} a_n$ is not viewed as the sum of infinitely many terms but as a limit of *finite* sums, $\sum_{n=1}^{N} a_n$; the apparent reference to infinity (∞) is explained away in finite terms. Likewise $\int_a^b f(x)\,dx$ is defined in terms of the

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

supremum and infimum of finite sums of areas; the apparent reference to infinitesimals (dx) is explained away in terms of finite intervals. This explaining away of apparent references to infinity in finite terms is characteristic of potential infinity. The above three equations are regarded as separate rules requiring separate proofs. The rigorous analytic theory of limits, infinite sums, differentiation and integration developed in the nineteenth century banished infinitesimals and infinite quantities in favour of arbitrarily small positive numbers ϵ and arbitrarily large natural numbers n. This certainly represents a supplanting of actual infinity by potential infinity. However, underlying this use of potential infinity are two uses of actual infinity: the concept of an infinite set (primarily the set of all natural numbers, $\mathbf{N}$, and the set of all real numbers, $\mathbf{R}$) and the concept of infinite quantifiers ('for all x, ...', 'there exists a δ ...').

Potential infinity and actual infinity are intertwined in modern mathematics. For example, we nowadays think of a line as something extending to infinity in both directions (unlike the ancient Greeks, who thought of a line as a line segment, indefinitely extendible in either direction). Nevertheless, each point on the line has a finite coordinate; no point has an infinite coordinate. The phrase 'going to ∞' refers to a direction, whereas 'going to 100' refers to a destination.

The idea of treating the infinite on the same footing as the finite, which is characteristic of actual infinity, is often very fruitful in mathematics. An obvious example is the points at infinity in projective geometry and complex analysis. But the same idea motivates concepts such as that of a Hilbert space. When one tries to generalise the theory of finite-dimensional vector spaces to infinitely many dimensions, one finds that many useful concepts and theorems break down. A possible reaction to this would be to say that infinite-dimensional vector spaces are fundamentally different from finite-dimensional ones, and so perhaps the general concept of vector space is not very useful. This would be hasty, however. It turns out that by imposing a few extra technical conditions (completeness and separability, and by considering closed linear subspaces rather than arbitrary linear subspaces), much of the finite-dimensional theory does generalise: for example, the notion of the closest point to a (closed) linear subspace, the idea of a basis-free treatment of the space, orthonormal expansions, the characterisation of a separable Hilbert space by its dimensionality, and the isomorphism between a Hilbert space and its dual space. It turns out that the distinction that matters is not the one between a finite-dimensional and an infinite-dimensional space (as we first thought) but the one between a Hilbert space (or possibly a separable Hilbert space) and other inner-product spaces. The mathematical pay-off of this is that it enables us to treat function spaces almost as if they were finite-dimensional.

Another example of this phenomenon is the concept of a compact set in topology. Finite point sets have many topological properties that do not hold for point sets in general. However, by formulating the notion of compactness the properties of finite sets can be generalised to a useful class of infinite sets. Hence there is no separate theory of finite sets in topology: it is simply subsumed in the theory of compact sets.

An advocate of actual infinity would draw a general moral from these examples: that the distinction between finite and infinite is less important than it naively seems, and often masks a deeper technical distinction; by recognising this deeper distinction we can bring the finite and the infinite under a single theory. Let us apply this thought to a philosophically controversial case: the notions of infinite set and infinite number seem very problematic at first sight, but, according to Cantor, set theory and arithmetic can be extended from the finite into the infinite by introducing the concept of the *transfinite*. The distinction between the transfinite and absolute infinity turns out to be more significant than the distinction between finite and infinite (see §2). Cantor's followers today therefore argue that acceptance of actual infinity is in the spirit of best mathematical practice. This viewpoint is typical of a platonist philosophy of mathematics.

On the other side, Brouwer's philosophy of intuitionism, and other schools of constructivism, view actual infinity with suspicion and attempt to found mathematics solely on potential infinity; they would regard this as in the spirit of best mathematical practice. Their views will be discussed in §3.

These questions of the nature of infinity in pure mathematics cannot be seen in isolation from applied mathematics and physics. It is a great defect of the literature on the paradoxes of mathematical infinity that it ignores the paradoxes of physical infinity, and vice versa. Mathematical and physical infinity are intimately interdependent in many philosophies of mathematics, including Field's [1980] nominalism, Hellman's [1989] modal structuralism, and logicism. Historically, ideas of infinitesimals and continuity arose from spatial and kinematic considerations. Even today, platonists frequently rely on infinitistic physical thought experiments to bolster their support of actual infinity. To many, it seems simply *obvious* that there could exist actual infinities in the physical world (for example, an actual infinity of stars), and they infer that actual infinity is also a coherent idea in pure mathematics [Russell, 1935–6; Benardete, 1964, 31]. I shall examine the idea of physical infinity in §4; it turns, out that it is so paradox-ridden that it counts more against actual infinity than in favour of it. However, it must be said that constructivists have so far failed to provide an alternative account of physics based on potential infinity.

In general, platonists believe that the same philosophical considerations apply to physical and mathematical infinity [Tait, 1986], whereas constructivists typically believe that different considerations apply [Dummett, 1994; Fletcher, 2002].

Whatever view you take of infinity, there is a pervasive problem you are certain to encounter, which I call the *horizon problem*. Imagine that you call yourself a 'finitist'; you proclaim that you only believe in finite things and that the infinite is illegitimate and incoherent. Someone asks you *how many* finite things you believe in. Embarrassingly, you have to admit that you believe in *infinitely many* of them. Notwithstanding all your 'finitist' rhetoric, you require a theory of infinity in order to give a global account of your mathematical worldview.

Suppose that, in the course of rising to this challenge, you come to accept that infinite collections (such as $\mathbf{N}$ and $\mathbf{R}$) are perfectly legitimate. Now, the same question recurs: *how many* infinite collections do you believe in? The answer is 'infinitely many', of course, but this latter infinity is a much larger, more unmanageable, kind of infinity than the humdrum infinities you do accept. In Cantorian terminology, the infinities you accept are called 'transfinite', and the infinity of everything transfinite is called 'absolute infinity'. To give a general account of transfinite infinity you require a theory of absolute infinity. You are worse off than when you were a finitist.

Suppose that, in desperation, you resolve to revert to finitism, and (to distance yourself as far as possible from infinity) to restrict yourself to the *feasible* finite, i.e., numbers you can actually count up to, such as 200, rather than practically inaccessible numbers such as 100^{100}. You resolve to found mathematics entirely on what is feasible. Again, the question recurs: how many feasible numbers are there? The answer, surely, is 'infeasibly many'.

All three of these positions (finitism, Cantorianism and feasibilism) suffer from the same problem: to characterise the domain of acceptable mathematical objects one has to use concepts that are not acceptable. This is the horizon problem, and it is the underlying problem of mathematical infinity; we shall encounter it repeatedly in §§2–3, though it is more acute in some theories than in others.

I had better admit my own bias here. I am a constructivist, and my views can be found in [Fletcher, 1998]. However, in this essay I shall try to expound and criticise each theory on its own terms, without trying to reach a final judgement between them.

2 ACTUAL INFINITY

2.1 Introduction

The 'actual' view of infinity goes naturally with a general philosophy of realism. By *realism* I mean the doctrine that the purpose of our physical and mathematical theories is to describe objective reality, that the meaning of statements is given by their truth conditions, that truth is independent of our means of knowing it, and that the proper names, variables, and other 'noun-like' expressions in statements denote *objects*. From such a standpoint it is natural to suppose that there is an infinity of objects and that quantifiers allow us to express infinitely many facts about the world in a single statement. Quantifiers were invented by Frege in 1879 and Peirce and Mitchell in 1883. Peirce and Mitchell devised the notation $\Sigma_i x_i$, meaning that x_i is true for *some* value of i, and $\Pi_i x_i$, meaning that x_i is true for *all* values of i, by a conscious analogy with infinite sums and infinite products in analysis [Peirce, 1885]. Nowadays we write these as $\exists i\ x_i$ and $\forall i\ x_i$, but we still (if we are realists) construe $\exists i\ x_i$ as a disjunction $x_1 \vee x_2 \vee x_3 \vee \cdots$ and $\forall i\ x_i$ as a conjunction $x_1 \wedge x_2 \wedge x_3 \wedge \cdots$ over all possible values of i. The fact that this disjunction and conjunction may contain infinitely many terms and so be impossible to compute is dismissed as irrelevant: the realist insists that the objects i exist independently of us, that the property x_i holds or fails to hold for each i irrespective of our ability to verify it, and that there is a *fact of the matter* about whether x_i is true for all, some or none of the objects i, irrespective of our ability to run through all the objects checking them.

An attractive feature of realism is that it gives us a uniform semantic account of both mathematical and nonmathematical discourse. As Benacerraf says, it allows us to give the grammatically similar sentences 'There are at least three large cities older than New York' and 'There are at least three perfect numbers greater than 17' a similar logical analysis [1973].

Nowadays, the realist view of mathematics is usually based on set theory. The choice of sets as the fundamental objects, rather than, say, functions, relations or categories, is a historical accident (e.g., functions are taken as the basic notion by [von Neumann, 1925], and [Bell, 1981] discusses the possibility of using categories); what is essential is that mathematics is seen as the study of a fixed universe of actually-infinite mathematical objects, existing independently of our ability to construct them or of any other anthropocentric considerations. A few set-theoretic realists believe that mathematics describes the same universe of objects as physical theories [Maddy, 1990; Mayberry, 2000]; but most regard mathematical objects as forming a separate realm disjoint from the physical world (this variety of realism is known as *platonism*).

This platonist orthodoxy is a comparatively recent phenomenon, dating back only a century or so. Before then, it was more common to view infinity as potential infinity. It is illuminating to look at how and why the change-over to actual infinity occurred. The transition arose out of the needs of nineteenth-century mathematics, particularly the arithmetisation of analysis. Four reasons can be traced.

1. *Rejection of spatial and temporal intuition.* Newton based his ideas of limits and differentiation on intuitions of motion; other mathematicians based their ideas of continuity on spatial intuition. These kinematic and geometric conceptions fell into disfavour in the nineteenth century, as they had failed to provide satisfactory theories of negative numbers, irrational numbers, imaginary numbers, power series, and differential and integral calculus [Bolzano, 1810, Preface]. Dedekind pointed out that simple irrational equations such as $\sqrt{2} \cdot \sqrt{3} = \sqrt{6}$ lacked rigorous proofs [1872, §6]. Even the legitimacy of the negative numbers was a matter of controversy in the eighteenth and nineteenth centuries [Ewald, 1996, vol. 1, 314–8, 336]. Moreover, Bolzano, Dedekind, Cantor, Frege and Russell all believed that spatial and temporal considerations were *extraneous* to arithmetic, which ought to be built on its own intrinsic foundations: 'it is an intolerable offence against *correct method* to derive truths of *pure* (or general) mathematics (i.e. arithmetic, algebra, analysis) from considerations which belong to a merely *applied* (or special) part, namely *geometry*' [Bolzano, 1817, Preface]. These considerations led them to reject the potential-infinity notion of a *quantity capable of augmentation without end*, with its connotations of time and change, and to replace it with the static notion of the *infinite set* of all possible values of the quantity.

2. *Quantifier combinations.* Before the nineteenth century, mathematical statements typically took the form of general laws ('$(x+y)^2 = x^2 + 2xy + y^2$') and constructions ('given any prime numbers we can construct a larger one'). These are expressible, in modern logical notation, using one or two quantifiers, and they can be understood readily in terms of potential infinity: they are all of the form 'given any numbers we can perform certain calculations, with certain results'. [Bolzano, 1817] pioneered the use of more logically complex statements requiring more quantifiers: for example, in modern notation, a sequence (x_n) is said to *converge* to a limit l iff

$$\forall \epsilon > 0 \, \exists N \, \forall n > N \; |x_n - l| < \epsilon;$$

   and a function $f : \mathbf{R} \to \mathbf{R}$ is said to be *continuous* iff

$$\forall a \, \forall \epsilon > 0 \, \exists \delta > 0 \, \forall x \; |x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon.$$

   Such statements are much harder to understand in terms of potential infinity. The quantified variables all vary in a complex interdependent way over an infinite range of values — it is hard to state this without thinking of the value-ranges as provided once and for all, as actually-infinite sets.

3. *Infinite sets as single objects.* The arithmetic theories of real numbers developed by Weierstrass, Cantor and Dedekind represented each individual real number by an infinite object (either a sequence of rationals or a Dedekind cut of rationals). Dedekind himself did not actually *identify* a real number with the corresponding Dedekind cut, but others did take this natural step [1888a]. The set of real numbers, $\mathbf{R}$, is then an infinite set of elements that are themselves essentially infinite. In the light of this it is hard to maintain the view that there is no such thing as actual infinity, only unboundedly varying finite quantities.

4. *Cantor's set theory.* Cantor was drawn into set theory by his investigations of Fourier series [Dauben, 1979, chapter 2]. He began by studying the representation of functions by trigonometric series, then he turned his attention to the sets of points on which such representations break down; this led him to study sets of points for their own sake, and then sets in general. For Cantor, a set was no longer simply the range of variation of a variable quantity; it had become a mathematical object in its own right.

Many forms of actual infinity are in use in mathematics today, for example, points at infinity in complex analysis and projective geometry, and infinitesimals in nonstandard analysis [Robinson, 1966] and smooth infinitesimal analysis [Kock, 1981; Bell, 1998]. Many others have been proposed and not received as much attention as they deserve, such as Peirce's theory of the continuum [Zink, 2001] and Vopěnka's alternative set theory [Vopěnka, 1979; Sochor, 1984]. Nevertheless, in this section I shall concentrate on the most important variety, Cantor's theory of transfinite sets and transfinite arithmetic. I shall restrict my attention to those aspects of set theory most relevant to infinity, neglecting other controversial aspects such as: the existence of sets and whether they are material or abstract objects; the existence of the empty set and singleton sets, and the comparison between set theory and mereology (for which see [Lewis, 1991]; the axiom of foundation and non-well-founded set theory [Aczel, 1988]; and the axiom of choice [Moore, 1982].

2.2 Infinite numbers

Cantor's boldest achievement was to extend the concept of number into the infinite, with his theory of transfinite ordinal numbers and transfinite cardinal numbers. Let us take the idea of cardinal number and try to isolate what is distinctive about Cantor's contribution.

There are two natural criteria by which one could compare the 'size' or 'multiplicity' of two classes, A and B (I am using the word 'class' in a rough, pre-Cantorian sense for the moment).

1. There are fewer As than Bs if every A is a B but not every B is an A.
2. There are as many As as Bs if there exists a bijection between the As and the Bs (i.e., a one-to-one correspondence, or a one-to-one mapping of all the As onto all the Bs).

Both these criteria were familiar since antiquity and were generally accepted for finite classes. But it was well known that they came into conflict when one tried to apply them to infinite classes, and this was sometimes used as an argument against infinite numbers. Leibniz said:

> The number of all numbers implies a contradiction, which I show thus: To any number there is a corresponding number equal to its double. Therefore the number of all numbers is not greater than the number of even numbers, i.e. the whole is not greater than its part. (Quoted in [Benardete, 1964, 44].)

The key step in resolving this conflict is to *separate* the two criteria. Maybe they are both valid, but they refer to different concepts of size. If we develop the consequences of the two criteria separately we might arrive at two valid theories of size, both applicable to infinite classes.

Thus Bolzano explicitly adopted (1) as his criterion of size and rejected (2) in the case of infinite classes [1851, §§19–24]. Bolzano's paper is notable for upholding the existence of actual infinity and for insisting that one infinity could be larger than another (for example, a line, bounded in one direction and unbounded in the other, can contain another such line (§19)). His rejection of criterion (2) led him into extraordinary complications when trying to evaluate sums of infinite series (§18). Nevertheless, his criterion of size has been incorporated into modern set theory in the form of the proper subset relation, and whenever we speak of, e.g., 'a maximal orthonormal set' or a 'maximal connected subset' we are using the word 'maximal' in Bolzano's sense.

Cantor, of course, adopted criterion (2) and replaced (1) with a weaker version.

> ($1'$) There are fewer or as many As as Bs if every A is a B.

By combining the two ideas ($1'$) and (2) we obtain the modern concept of the *cardinality*, $|A|$, of a class A:

$$\begin{array}{ll} |A| = |B| & \text{iff there is a bijection between } A \text{ and } B; \\ |A| \leq |B| & \text{iff there is a bijection between } A \text{ and a subclass of } B; \\ |A| < |B| & \text{iff } |A| \leq |B| \text{ but } |A| \neq |B|. \end{array}$$

If $|A| = |B|$ then A and B are said to have the same *power*, or the same *cardinal number*. Beginning with his [1874] paper, Cantor accomplished cardinal comparisons between the set of natural numbers ($\mathbf{N}$), the set of real numbers ($\mathbf{R}$), the set of algebraic numbers, the set of irrational numbers, the one-dimensional interval $[0, 1]$, and the two-dimensional square $[0, 1]^2$. He showed that there is an unending succession of infinite cardinal numbers, of which the smallest is $|\mathbf{N}|$, known as *countable* infinity. In parallel with this, Cantor developed a theory of transfinite *ordinal* numbers, $0, 1, 2, \ldots \omega, \omega{+}1, \omega{+}2, \ldots \omega2, \omega2{+}1, \omega2{+}2, \ldots$, thinking of them initially as steps in a transfinite iteration and later as order-types of well-ordered sets. A *well-ordered set* is a set equipped with an ordering relation under which every nonempty subset has a least element. Two well-ordered sets are said to be *similar*, or *of the same order-type*, iff there exists a similarity between them (a similarity is an order-preserving bijection). Similarities are the measuring stick for ordinal numbers, just as bijections are the measuring stick for cardinal numbers.

Cantor defined operations of addition, subtraction, multiplication, division and exponentiation for ordinal numbers and investigated their unique prime factorisation. He also defined addition, multiplication and exponentiation for cardinal numbers, and used ordinal numbers to index the sequence of cardinal numbers: the sequence of infinite cardinal numbers could be listed as $\aleph_0, \aleph_1, \aleph_2, \ldots, \aleph_\omega, \aleph_{\omega+1}, \ldots$, with every infinite cardinal number occurring in the list as $\aleph_\alpha$ for a unique ordinal number α. He developed a distinctive concept of *set*, under which every set had a cardinal number and every well-ordered set had an ordinal number.

These achievements made an immense impression on Cantor's successors. By the simple expedients of taking actual infinity seriously and adopting bijections and similarities as measures of size, Cantor had replaced the mediaeval metaphysical morass of infinity with determinate mathematics. The fact that one can do *calculations* with infinities, one can show that $5(\omega{+}1)(\omega^2 3{+}4) = \omega^3 3{+}\omega 4{+}5$ and $\aleph_{37} \times \aleph_{20} = \aleph_{37}$, convinced mathematicians that transfinite numbers were just as real as finite numbers and that Cantor's theory represented a genuine advance over all previous thinking about infinity. Mathematicians have a strong sense of the *reality* and *concreteness* of anything they can calculate with. A few mathematicians

(Kronecker, Baire, Brouwer, Poincaré, Weyl) resisted the move to actual infinity; the great majority applauded it. (However, it might be argued that if they had known at the time how intractable the evaluation of cardinal exponentials such as $2^{\aleph_0}$ was they might have been less willing to embrace transfinite arithmetic.)

2.3 The paradoxes of set theory

Even as set theory was finding its feet and actual infinity was establishing itself as mathematical orthodoxy, serious contradictions were discovered at the heart of set theory, which were seen at the time as a 'crisis' in the foundations of mathematics. Set theorists at this time (the beginning of the twentieth century) assumed a *comprehension principle*:

$$\exists S\,\forall x\ (x \in S \Leftrightarrow P(x))$$

which says that, given a well-defined property P applicable to objects of any kind, there exists a set S consisting of all the objects x for which $P(x)$ holds. The motivation for this was the commonly held belief that consistency implies existence. The view was that, if one defines a mathematical system by specifying the types of elements it contains, the relations and functions that apply within the system, and the axioms that characterise them, and if one does so in a rigorous, unambiguous, consistent and complete manner (that is, if the axioms determine unambiguously whether each relation holds between any two elements and the value of each function on any argument), then the system is guaranteed to exist. (This doctrine is nowadays known as 'plenitudinous platonism' [Balaguer, 1998].) Now, a set is a mathematical system of a very simple kind, having only one relation, $\in$; and so if we specify unambiguously what it means for an arbitrary object x to belong to the set then we have done enough to guarantee the set's existence. This is the justification for the comprehension principle.

This belief that consistency implies existence arose from the successes of nineteenth-century mathematics in generalising algebra to ever-larger number systems and establishing Euclidean and non-Euclidean geometry on a firm axiomatic footing. Dedekind believed that mathematical objects were 'free creations of the human mind', that human beings possess 'divine' powers to create any well-defined mathematical system [1888a, b]. Cantor believed that any consistently and precisely defined mathematical system was guaranteed to occur in the external world [1883, §8]. Poincaré [1905] and Hilbert [1900] believed that, in mathematics, existence simply *is* consistency. Frege provided an explicit articulation of this idea, in the form of his 'context principle' [1884, §60; 1919], which states that the meaning of a word consists in the contribution it makes to the meanings of the sentences in which it may occur; if this contribution can be specified adequately then the word is guaranteed to have reference. He applied the context principle in his *Grundgesetze* [1893, §§3,10] to fix the meaning of the course-of-values of a function; he argued that to do this it sufficed to provide an equality criterion for courses-of-values; this was his Basic Law V, his equivalent of the comprehension principle.

The set-theoretic paradoxes arise from the comprehension principle. Burali-Forti's paradox appeared in 1897 (in a disguised form). Let P be the property of being an ordinal number; S is therefore the set of all ordinal numbers. Then it is clear that S is well-ordered, by the usual ordering of ordinals. Every well-ordered set has an ordinal number as its order-type, so let Ω be the order-type of S. Now, Ω must be greater than all the ordinal numbers in S; but this contradicts the fact that S contains all ordinal numbers.

Around 1900–2, Zermelo and Russell independently discovered what is generally known as 'Russell's paradox'. Let P be the property that holds of x iff x is a set and $x \notin x$. Then S is the set of all sets that do not belong to themselves. The definition immediately gives $S \in S \Leftrightarrow S \notin S$, a logical contradiction (even in intuitionistic logic).

'Cantor's paradox' (essentially the argument in [Cantor, 1899c]) arises by taking P as the property of being a set. Thus S is the set of all sets. Let $\mathcal{P}(S)$ be the power set of S, i.e., the set of all subsets of S. Then S is smaller in cardinality than $\mathcal{P}(S)$, by Cantor's diagonal argument [1891]; but this contradicts the obvious fact that $\mathcal{P}(S) \subseteq S$.

After much confusion and controversy, it came to be generally agreed that the fallacy in each of these arguments was the application of the comprehension principle: there is no set of all ordinal numbers, no set of all sets that are not members of themselves, and no set of all sets. It was concluded that the comprehension principle was unsound, at least in its full generality. Yet some principle of this sort is necessary if we are ever to be able to claim the existence of any set or other mathematical system. So, lacking a clear understanding of what was wrong with the comprehension principle, set theorists took the line of least resistance and assembled lists of weaker principles that would allow them to carry out all the normal operations of mathematics without reproducing the three paradoxes. This led to an axiomatic theory called ZFC (Zermelo-Fraenkel set theory with the axiom of choice), which is nowadays considered the standard version of set theory. The axioms are as follows.

AXIOMS FOR BASIC SETS

Axiom of the empty set ($\emptyset$ exists): $\exists y \forall z\ z \notin y$

Axiom of infinity (there exists an infinite set): $\exists y\ (\emptyset \in y \wedge \forall z \in y\ z \cup \{z\} \in y)$

SET-BUILDING AXIOMS

Axiom of pairing ($\{x_1, x_2\}$ exists): $\forall x_1, x_2 \exists y \forall z\ (z \in y \Leftrightarrow (z = x_1 \vee z = x_2))$

Axiom of union ($\bigcup x$ exists): $\forall x \exists y \forall z\ (z \in y \Leftrightarrow \exists w \in x\ z \in w)$

Axiom of power set ($\mathcal{P}(x)$ exists): $\forall x \exists y \forall z\ (z \in y \Leftrightarrow z \subseteq x)$

Axiom of separation ($\{z \in x \mid P(z)\}$ exists): $\forall x \exists y \forall z\ (z \in y \Leftrightarrow (z \in x \wedge P(z)))$

Axiom of replacement ($\{F(u) \mid u \in x\}$ exists): $\forall x \exists y \forall z\ (z \in y \Leftrightarrow \exists u \in x\ z = F(u))$

GENERAL AXIOMS ABOUT SETS

Axiom of extensionality ($x = y$ iff x and y have the same elements):
$\forall x, y\ (x = y \Leftrightarrow \forall z\ (z \in x \Leftrightarrow z \in y))$

Axiom of foundation (there are no infinite membership chains $a_0 \ni a_1 \ni a_2 \ni \cdots$):
$\forall x \neq \emptyset\ \exists y \in x\ x \cap y = \emptyset$

Axiom of choice (for any set x of nonempty disjoint sets, there is a set w containing *one* element from each set in x):
$\forall x\ ([\forall y, z \in x\ y = z \Leftrightarrow y \cap z \neq \emptyset] \Rightarrow \exists w \subseteq \bigcup x\, \forall y \in x\, \exists u\ w \cap y = \{u\})$

(In the axiom of separation P is any property of sets; in the axiom of replacement F is any operation that transforms sets to sets; for the sake of definiteness it is common to assume that P and F are expressible by means of formulae in the first-order language of ZFC.) Most of these axioms were introduced by Zermelo [1908]; the axioms of replacement and foundation were added later by others.

I have grouped the axioms into three classes to suggest a quasi-constructive way of 'generating' the ZFC universe of sets. We start with the basic sets (the empty set and the given infinite set), and apply the set-building axioms repeatedly to construct further sets. All the set-building axioms are of the form 'given a set x (or two sets x_1, x_2) we can construct another set y from it'. The construction process is subject to the conditions imposed by the 'general' axioms: two sets are considered equal iff they have the same elements; no set may involve an infinite regress or a vicious circle of membership; and it is always possible to make simultaneous choices of elements from any number of nonempty sets. This quasi-constructive view is known as the *iterative conception of sets*; it will be discussed further in §2.5.

2.4 Cantor and absolute infinity

The paradoxes of set theory were less paradoxical to Cantor than to his contemporaries. When Cantor was made aware of the paradoxes he was able to point out that he had anticipated something of this sort and made allowance for it in his [1883] paper. Indeed, he had drawn a distinction between the *transfinite* and *absolute infinity*. Infinite sets are transfinite: they 'can be determined by well-defined and distinguishable *numbers*' [1883, §5]. The sequence of all ordinal numbers is absolutely infinite, and exceeds all rational grasp: 'the absolute can only be acknowledged but never known — and not even approximately known' [1883, endnote 2]. In later years he expressed this in terms of a distinction between 'consistent multiplicities' (of transfinite size) and 'inconsistent multiplicities' (of absolutely infinite size). As I have already said, Cantor shared the widespread assumption that any consistent mathematical system exists; but he had a novel conception of consistency. He believed that for a system to be consistent it is necessary not merely that its elements and their mutual relations be consistently specified but that the elements be able to coexist consistently as a whole.

> If we start from the notion of a definite multiplicity (a system, a totality) of things, it is necessary, as I discovered, to distinguish two kinds of multiplicities (by this I always mean *definite* multiplicities).
>
> For a multiplicity can be such that the assumption that *all* of its elements 'are together' leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as 'one finished thing'. Such multiplicities I call *absolutely infinite* or *inconsistent multiplicities*.
>
> As we can readily see, the 'totality of everything thinkable', for example, is such a multiplicity; later still other examples will turn up.
>
> If on the other hand the totality of the elements of a multiplicity can be thought of without contradiction as 'being together', so that they can be gathered together into '*one* thing', I call it a *consistent multiplicity* or a 'set'. [Cantor, 1899a]

> *Thus there exist* determinate multiplicities that are *not also unities* — i.e. multiplicities such that a real 'coexistence of all their elements' is *im*possible. These are the ones I call 'inconsistent systems'; the others I call 'sets'. [Cantor, 1899c]

It sounds from this as if a set is a special kind of multiplicity; to understand sets we first need a theory of multiplicities, then we need a dynamical theory explaining how elements are 'gathered' into multiplicities and how they sometimes form stable wholes and sometimes fall apart under the mutual repulsion of their elements (like an unstable atomic nucleus). But this, I think, would be to take the talk of multiplicities and gathering too literally. 'Multiplicities', 'systems' and 'totalities' are merely figures of speech; sets are the only kind of collective object that Cantor believes in. A set is simply a single object containing other objects. When Cantor says that 'The system Ω of all [ordinal] numbers is an inconsistent, absolutely infinite multiplicity', he is saying that the ordinal numbers do not form any sort of collective object at all: they are just too numerous.

Thus Cantor replaced the traditional distinction between the finite and the infinite by a distinction between the finite, the transfinite and the absolutely infinite. The transfinite includes all the infinite sets commonly used in mathematics (such as **N**, **R** and separable Hilbert spaces), but it also shares many of the properties of the finite: it is numerically determinate, humanly graspable, 'limited', 'increasable' and 'finished'. The absolutely infinite is beyond all rational determination and is not itself an object of mathematical study. In Cantorian set theory the transfinite/absolute distinction assumes central importance and plays a similar role to that traditionally played by the finite/infinite distinction; whereas the finite/transfinite distinction fades into relative unimportance. The theory of cardinal and ordinal arithmetic is developed in a uniform way for all sets, and the finite part of it emerges merely as a special case, not receiving a different foundational treatment from the transfinite part (Hallett calls this point of view 'Cantorian finitism' [1984]). Mayberry emphasises the point further by applying the word ‘finite’ to all finite and transfinite sets [2000]. In Mayberry’s terminology, Cantor did not develop a theory of infinite sets but *extended the finite* to cover sets such as **N** and **R** that had traditionally been conceived as infinite. (I shall not follow this terminology, however, as I wish to retain the use of the word ‘finite’ in the usual sense of ‘smaller than $\aleph_0$’.)

It is clear from this how fundamental Cantor’s distinction between the transfinite and absolute infinity is to his project; it is not just a device for avoiding the paradoxes. But how soundly based is the distinction? In particular, what prevents us from treating an absolutely infinite ‘multiplicity’ as a definite object? Cantor does after all describe an absolutely infinite multiplicity as ‘fully-determinate, well-defined’ [1899c], meaning I think that it has a well-defined criterion for membership. But he believes it is not *numerically* determinate, meaning that it is too big to have a cardinality. Yet an obvious objection arises: why can we not compare absolutely infinite multiplicities with respect to cardinality, just as we do with transfinite sets? Cantor showed that there was a bijection $\alpha \mapsto \aleph_\alpha$ between the ordinal numbers and the infinite cardinal numbers; does this not establish that the multiplicity of ordinal numbers is of the same size as the multiplicity of infinite cardinal numbers? Moreover, in modern set theory with the global axiom of choice we can show that all absolutely infinite multiplicities are of equal cardinality to V, the multiplicity of all sets. Hence there seems to be a well-defined cardinality common to all absolutely infinite multiplicities.

Now, Cantor had defended the introduction of actual infinity against the objections of Aristotle by saying that ‘determinate countings can be carried out just as well for infinite sets as for finite ones, provided that one gives the sets a determinate law that turns them into *well-ordered* sets’ [1883, §4]. In other words, the fact that one can carry out cardinal and ordinal comparisons in a mathematically crisp and unambiguous way, free of any arbitrariness, demonstrates the validity of infinite sets and transfinite arithmetic. But the very same considerations suggest that there is a determinate cardinality of all sets and an order-type of all ordinal numbers, which implies a higher realm of ‘super-sets’ of absolutely infinite size. A Cantorian must reject this conclusion, for it subverts the whole idea of absolute infinity as exceeding all mathematical determination. Yet the Cantorian seems to be without resources to resist the argument.

Dummett sets out the difficulty in the following terms. The idea of a transfinite number seems self-contradictory when one first meets it, he says, as one is used to thinking of a number as something that can be arrived at by counting. However, the beginner can be persuaded that it makes sense to speak of infinite sets of different sizes.

> When he [the beginner] has become accustomed to this idea, he is extremely likely to ask, “How many transfinite cardinals are there?” How should he be answered? He is very likely to be answered by being told, “You must not ask that question.” But why should he not? If it was, after all, all right to ask, “How many numbers are there”, in the sense in which “number” meant “finite cardinal”, how can it be wrong

> to ask the same question when "number" means "finite or transfinite cardinal"? A mere prohibition leaves the matter a mystery. It gives no help to say that there are some totalities so large that no number can be assigned to them. We can gain some grasp of the idea of a totality too big to be counted, even at the stage when we think that, if it cannot be counted, it does not have a number; but, once we have accepted that totalities too big to be counted may yet have numbers, the idea of one too big even to have a number conveys nothing at all. And merely to say, "If you persist in talking about the number of all cardinal numbers, you will run into contradiction" is to wield the big stick, not to offer an explanation. [Dummett, 1994]

Clearly we need a further account of the difference between consistent and inconsistent multiplicities to extricate ourselves from this difficulty. I shall turn to this in the next subsection.

2.5 Three views of sets

We are seeking an account of how a 'multiplicity' of objects can be joined together into a single object, a 'set'. There are three main approaches:

- sets as *classes*;
- the *limitation of size* view;
- the *iterative conception of sets*.

By a *class* I mean a property, viewed extensionally. For example, the property of being a subset of $\mathbf{R}$ is (let us suppose) fully determinate: it is determined which objects possess the property and which don't. We can say that x belongs to the class $\mathcal{P}(\mathbf{R})$ iff x satisfies this property; talk of classes is just an extensional idiom for talk of satisfying properties. We refer to two classes as 'equal' iff they have the same elements, even though the two properties may be different. By forming the property 'subset of $\mathbf{R}$' we have enabled ourselves to refer, in a single breath, to infinitely many things (all the subsets of $\mathbf{R}$); we have gathered them, mentally, into a unity, which is the essential requirement for a set.

But what is a property? Ambiguity on this point vitiated Russell's account of the paradoxes in terms of 'propositional functions' [1906]. If our explanation of sets in terms of properties is to achieve any real reduction, we must understand properties syntactically, as given by sentences containing a single free variable, e.g., 'x is a subset of $\mathbf{R}$'. There is no restriction on the language; any meaningful declarative sentence with one free variable is considered to denote a property. This was Frege's notion of a *concept* [1891].

How do we escape the paradoxes, on this view? The properties expressed by the sentences 'x is an ordinal number', 'x does not belong to x', and 'x is a class' must somehow be rejected. It is not plausible to say that the sentences are not meaningful: given a particular object, we *do* know what it means to say that it is an ordinal number, or that it does not belong to itself, or that it is a class. We could, however, say, with Russell, that 'a statement about x cannot in general be analysed into two parts, x and what is said about x' [1906]. That is, if one takes a sentence of the form 'A does not belong to A', and tries to analyse it into a part A and a schema 'x does not belong to x', this works syntactically (i.e., the sentence is recoverable by substituting A back into the schema), but it somehow fails semantically. The proposition expressed by the sentence just cannot be carved up that way. This is all rather mysterious.

A more hopeful line is to assimilate the set-theoretic paradoxes to the semantic paradoxes. The semantic paradoxes are diagonalisation arguments involving sentences and properties, which work in a similar self-referential way to the set-theoretic paradoxes. I shall just outline the three main semantic paradoxes.

- (Grelling's paradox) Call a property *heterological* iff it does not satisfy itself. Is heterological itself heterological?
- (Berry's paradox) Let N be the least natural number not definable in English in less than one hundred words. Then N is definable in English in less than one hundred words (we have just done so in the previous sentence!).
- (Richard's paradox) Let E be the set of every real number whose decimal expansion is definable in English. Enumerate the elements of E by lexicographic ordering of their English definitions. By diagonalisation on the decimal expansions we can construct a real number not in E: we have thus defined an indefinable number.

These paradoxes need to be resolved by tightening up our language in some way; either that or we need to live with the fact that our language is powerful enough to generate nonsense. Whatever resolution we adopt for the semantic paradoxes is likely also to cover the set-theoretic paradoxes (construed in terms of properties). Hence the mere presence of the paradoxes, troublesome though they are, is not an objection to the class view of sets.

A set theory based on this approach would take account of Poincaré's [1906b, §IX; 1910; 1913, chapter IV] and Weyl's [1921, 1925–7] considerations of the 'vicious circle' principle and 'predicative' definitions. It would probably be similar to Whitehead & Russell's type theory [1910] or Quine's set theory [1937].

The *limitation of size* view rejects any such intensional origin of sets. A set is defined as a multiplicity or plurality of objects that is *quantitatively determinate* and hence may be thought of as a single thing. There is no doubt that this is in line with Cantor's own thinking: see especially his critique of Frege [Cantor, 1885]. On this view, the difference between consistent and inconsistent multiplicities is that the latter are simply too big to have a cardinality. Neither Cantor nor any of his successors has offered any explanation of why cardinality should be the decisive factor. Indeed, there are alternative versions of set theory in which a set can have a non-set as a subclass [Vopěnka, 1979; Sochor, 1984].

Mayberry [2000] gives a lucid and vigorous modern presentation of set theory from the 'limitation of size' viewpoint. He addresses a *logical* question that twentieth-century set theorists were very slow to take up: given that absolute infinity 'can only be acknowledged but never known', how can it be legitimate to quantify over all sets? The difference between transfinite and absolutely infinite multiplicities is so fundamental that one would expect it to be reflected in a different semantics and logic for 'bounded' quantifiers ($\forall x \in S$, $\exists x \in S$, ranging over a set) than for 'unbounded' quantifiers ($\forall x$, $\exists x$, ranging over the whole universe of sets). Yet ZFC set theory allows unbounded quantifiers to be used freely, and combined with the propositional connectives in arbitrary ways, without any restrictions or precautions, as if the set universe were a domain like any other.

Mayberry gives the following reason for being suspicious of unbounded quantification (§§3.5, 7.2). If unbounded quantification were allowed without restriction then we could use it to define an identity criterion for classes of sets (or 'species', as Mayberry calls classes):

$$S = T \quad \text{iff} \quad \forall x\ (x \in S \Leftrightarrow x \in T).$$

If unbounded quantifiers were subject to classical logic then any classes S and T would determinately be either equal or unequal. Given this firm criterion of identity, it would be hard to resist the conclusion that classes were *objects*, and from them we could form pluralities of classes, and pluralities of pluralities of classes, and so on. This subverts the Cantorian definition of 'set': a set was defined as a multiplicity that is also a unity (an object), so we cannot admit any other kind of collective object. Hence a Cantorian *must* restrict unbounded quantification in some way. Mayberry discusses various possible theories involving first- or second-order quantification over sets or classes, and concludes that it is better to allow only Π_1 and Σ_1 propositions (i.e., $\forall x \cdots \forall y\ A$ and $\exists x \cdots \exists y\ A$, where the initial quantifiers range over all sets and A contains only bounded quantifiers), and to give them a quasi-intuitionistic semantics and logic.

Aside from this logical question, there is a persistent doubt (first raised by [Russell, 1906]) about the meaningfulness of the fundamental notion of 'limited in size'. What is it exactly that determines whether a multiplicity is too large to have a cardinality? How big is too big?

Imagine a mathematician who whole-heartedly accepts the thesis that mathematics is based on sets, that a set is a plurality of limited size, and that size is measured by bijections, but who nevertheless refuses to accept set theory in its conventional form, preferring instead to adopt one of the following positions.

> *Position 1:* all sets are finite (in the usual sense of being smaller than $\aleph_0$); their cardinalities can be measured by natural numbers; absolute infinity coincides with infinity.
>
> *Position 2:* only finite and countably infinite sets exist; the process of constructing ordinal numbers can never reach ω_1 (the first uncountable ordinal number); ω_1 represents absolute infinity.

> *Position 3:* finite and countably infinity sets exist, and the possibility of uncountable sets is left open, but the power-set axiom is rejected; the power 'set' of any infinite set is absolutely infinite.

The problem here is that all three unorthodox positions are compatible with Cantor's thesis of limitation of size and they can all be well motivated.

As regards position 1, it is generally accepted that the existence of infinite sets is merely a hypothesis, albeit a very useful one for mathematics. Mayberry develops two alternative set theories, called 'Cantorian' and 'Euclidean', based on affirming and denying that a countable infinity is 'limited', respectively [2000]. Cantor shrugs off the question with the comment that we cannot even prove that *finite* multiplicities are sets [1899b]. His casualness on this point is disconcerting. Is it really the case that we cannot tell whether $\{a, b, c, d, e\}$ is absolutely infinite? If so, is that because the relevant facts are somehow hidden from us, or because the concept of limited size is, after all, meaningless?

Pollard suggests the following criterion: 'things are limited in number whenever they are less numerous than other things'; in other words, a plurality is limited iff it is *increasable* [1996]. By Cantor's diagonal argument, the natural numbers are less numerous than the sets of natural numbers. Hence the plurality of natural numbers is limited. The fallacy of this argument is that it assumes the two pluralities involved *have* sizes, i.e., that they are sets. The diagonal argument shows that, *if* the plurality of natural numbers and the plurality of sets of natural numbers have sizes, then the former size is less than the latter size. This is no help in establishing that any infinite plurality has a size.

Position 2 is naturally suggested by reading Cantor's account of the generation of the ordinal numbers [1883, §§1,11]. Cantor considered the ordinals as produced in a quasi-constructive way by applying two generating principles repeatedly, starting from 0:

1. given an ordinal α we can form its successor $\alpha + 1$;
2. given any succession of ordinals $\alpha, \beta, \ldots$ already generated, with no greatest member, we can form the *limit* of them, the first ordinal greater than all of them.

This process, known as transfinite iteration, generates the ordinals

$$0, 1, 2, 3, \ldots, \omega, \omega{+}1, \omega{+}2, \ldots, \omega2, \omega2{+}1, \omega2{+}2, \ldots, \omega3, \omega3{+}1, \ldots, \omega4, \ldots, \omega^2, \omega^2{+}1, \ldots$$

where $\omega, \omega2, \omega3, \omega4, \omega^2$ are generated by the second principle and the others shown (apart from 0) are generated by the first principle.

It seems clear that by applying the two generating principles we can form a long (but always countable) succession of countable ordinal numbers. But, as Lake asks, how are we ever to reach the uncountable ordinals [1979, §1.3]? It seems that we would need to generate uncountably many countable ordinals to justify the leap to ω_1; but the possibility of doing this is itself in question. How can we pull ourselves up by our countable bootstraps into the realm of the uncountable? Cantor tries to accomplish this step by defining the second number class, (II), as the aggregate of all countable ordinal numbers formable by the two generating principles; he says that we can continue from here to form further number classes of ordinal numbers [1883, §12], so he seems to believe that in forming (II) we have thereby formed ω_1. (Cantor appeals to a third principle, a 'restricting or limiting principle', whose function seems to be to help us in delimiting the number classes.) But if it is acceptable to generate ω_1 as the limit of uncountably many countable ordinals, why is it not acceptable to generate Ω, the absolutely infinite order-type of all transfinite ordinals, as the limit of the succession of all transfinite ordinals? It would be extremely natural to identify ω_1 with Ω and to regard Cantor's generating process as inherently limited to the countable.

Position 3 is naturally suggested by Cantor's diagonal argument. It is conventional to say that the diagonal argument shows that the power set of any set S is strictly larger than S. But this formulation is question-begging, as I pointed out above. The conclusion of the diagonal argument is really that, given any set A of subsets of S, with $|A| = |S|$, we can always construct another subset of S outside A. It does not establish that there is a set of *all* subsets of S. If one couples the diagonal argument with Russell's, Poincaré's and Weyl's arguments about the vicious circle principle and impredicativity, one may easily be led to the conclusion that $\mathcal{P}(\mathbf{N})$ is an inexhaustible multiplicity of sets, so big that when we try to grasp it in full we always fall short. In short, $\mathcal{P}(\mathbf{N})$ is an excellent candidate for absolute infinity. (Compare this with the argument of [Cohen, 1966, 151].)

Power sets were used implicitly by Cantor [1891]; the axiom asserting the existence of power sets was first stated by [Zermelo, 1908], without any justification. Lake asserts the axiom as obvious:

> The sum and power set axioms follow as it is inconceivable that an inconsistent multiplicity could be obtained from a set by one of these visualizable operations. This is even clearer if we assume that all inconsistent multiplicities are the same size, for then the power set axiom, for instance, says that there is no set for which the collection of all its subcollections is the same size as the Absolute. [Lake, 1979]

Mayberry [1994] regards the power-set axiom as 'obviously true', from a limitation of size viewpoint. In [2000] he is more tentative, describing it as 'the most powerful, and least self-evident, of the finiteness principles' (p. 117), but still 'highly plausible' (p. 123) (recall that Mayberry uses the word 'finite' to mean limited in size).

> The subsets of a plurality of determinate size are simply "there", in whatever multitude, definite or indefinite, they may compose. Surely it is implausible to suppose that multitude to be absolutely infinite. Surely we cannot conceive that the absolutely infinite could be rooted in, or could emerge from, a particular instance of the finite in such a manner. (p. 124)

Yet set theorists do believe that the absolutely infinite universe of sets emerges from nothing (or from two or more urelements in Mayberry's case), in the manner described by the cumulative hierarchy of sets (see below for a definition of the cumulative hierarchy). This is just as astonishing and counter-intuitive, when one first meets it, as the idea that a set could have an absolute infinity of subsets. We are all so *familiar* with ZFC set theory that we become insensible to its bizarre aspects and we tend to mistake familiarity for self-evidence. We have a vivid diagram in our minds of the cumulative hierarchy, in which $\mathcal{P}(S)$ lives one level above S, while absolute infinity is represented by an arrow pointing up towards heaven. This diagram is so entrenched in our thinking that it obstructs us from considering with an open mind the possibility that $\mathcal{P}(S)$ is absolutely infinite. Hallett [1984, chapter 5] examines in detail the attempts to justify the power-set axiom from a limitation of size viewpoint and concludes that they cannot succeed.

The problem revealed by each of these three unorthodox positions I have sketched is that our notion of limitation of size seems insufficient to justify the axioms supposedly based on it. The axioms that I called 'set-building axioms' in §2.3, together with the axiom of infinity, are, from the limitation of size viewpoint, intended to articulate the meaning of 'limited size'. Indeed, one would hope that the meaning of 'limited size' would be clarified in the course of arguing for the axioms (one usually understands a mathematical concept better when one has seen it used in arguments). And indeed it can be said that the separation and replacement axioms arise naturally from the fundamental principles of cardinality, $(1')$ and (2), with which we began in §2.2. However, the same cannot be said for the other axioms, particularly the power-set and infinity axioms (and possibly also the union axiom); the justifications given by set theorists for these axioms have a circular, or even vacuous, character; one could reject power sets or infinite sets and still claim to be true to the principle of limitation of size. I am suspicious of mathematical concepts that do no work; it is the axioms that do the work in set theory, not the principle of limitation of size, and there seems to be an unbridgeable gap between the principle and the power-set and infinity axioms. These two axioms are central to Cantor's project, for we need them to show the existence of infinite sets of different sizes.

Let us turn to the third approach to set theory, the *iterative* conception of sets. As we have seen, Cantor viewed the ordinal numbers as generated by a transfinite iteration, using his two principles of generation. A similar transfinite iteration process can be used to generate the entire universe of sets. We think of sets as generated in *stages*, indexed by ordinal numbers. The collection of sets formed at or before stage α is called V_α (where α is any ordinal number), and is defined by

(i) $V_0 = \emptyset$,

(ii) $V_{\alpha+1} = \mathcal{P}(V_\alpha)$,

(iii) if α is the limit of a succession of lesser ordinals then $V_\alpha = \bigcup_{\beta<\alpha} V_\beta$.

Clause (i) says that we start with no sets. Clause (ii) is analogous to Cantor's first principle of generation: at stage $\alpha + 1$ we form all sets consisting of sets available at stage α. Clause (iii) is analogous to Cantor's second principle of generation: if α is formed by taking the limit of a succession of pre-existing ordinals, then no new sets are formed at stage α; the only sets available are those that were already available at some previous stage.

It is a consequence of this definition that if $\beta < \alpha$ then $V_\beta \subset V_\alpha$; hence this procedure gives a *cumulative hierarchy* of sets. Each set is formed at a certain stage α; it will therefore belong to V_α and all V_γ for $\gamma > \alpha$. Indeed, a set is *defined* as something produced by this process at some stage. Notice that Cantor's 'inconsistent multiplicities' are never produced at any stage.

This iterative conception of sets emerged from the work of [Mirimanoff, 1917] and [Zermelo, 1930], and is followed by [Gödel, 1983; Boolos, 1971; Wang, 1974; Shoenfield, 1977] (1971) (1974) (1977) and [Maddy, 1990]. Popular though the iterative conception is today, many authors warn us in footnotes against taking it too literally. Maddy says 'Of course, the temporal and constructive imagery is only metaphorical; sets are understood as objective entities, existing in their own right' (p. 102). Likewise Shoenfield says 'We should interpret "before" here in a logical rather than a temporal sense' (p. 323). The point is that a set formed at stage α is logically dependent on sets formed at lesser stages.

Cantor himself rarely considered sets of sets, so there is no sign of the cumulative hierarchy, whether construed iteratively or not, in his writings. His position on temporal imagery was rather ambiguous: officially he was against it [Hallett, 1984, 28; Cantor, 1885], yet he relied extensively on ideas of finite and transfinite iteration.

> By a *finite* set we understand a set M which arises out of an original element through the successive addition of elements in such a way that also the original element can be achieved *backwards* out of M through *successive removal* of elements in reverse order.

This quotation is from 1887–8 and is translated in Hallett [1984, 147]; Hallett goes on to point out that Cantor's proof that every infinite set has a countable subset relies on successively selecting an infinite sequence of elements from the set. Moreover, Cantor's belief that every set could be well-ordered seems to have depended on the idea that transfinite iterations could be carried out and completed [1899a]. Gödel also seems to take transfinite iteration literally. He points out that the kinds of set used in mathematics are always sets of integers, sets of real numbers, set of functions from $\mathbf{R}$ to $\mathbf{R}$, or sets of some other limited type, rather than sets in general; he proposes to take the operation 'set of' as primitive, and to build up the universe of sets by transfinite iteration of this operation.

The iterative conception of sets leads to similar results to the limitation of size conception, and many set theorists endorse both conceptions. Nevertheless, the two are not equivalent. Mayberry, a supporter of the limitation of size view, regards the notion of transfinite iteration (or even finite iteration) as incoherent and utterly unsuitable as a basis for mathematics [2000]; his book is useful for showing how the limitation of size view can be completely disentangled from the iterative conception.

It should be noticed, however, that the iterative conception does not provide *justifications* for the axioms; rather, the axioms are *built into* it. Shoenfield tries to justify the power-set axiom from the iterative conception [1977, 326], and likewise Maddy claims that the iterative conception provides 'the most satisfying account' of the power-set axiom [1997, 53]; see also [1988]). But in reality the iterative conception does not provide any account at all; it just takes the power-set operation for granted, when it assumes that we can collect together *all* sets consisting of objects generated so far. The axiom of infinity is also presupposed: it is assumed that given a countably infinite sequence of stages there is a limit stage beyond them. In Wang's and Shoenfield's versions, the axiom of replacement is also presupposed in their assumptions about stages; they seem to be relying on an idea of limitation of size here, so I am inclined to agree with Boolos that the axiom of replacement is not intrinsic to the iterative conception. Maddy [1997] relies more on 'extrinsic' justifications for the axioms than on any conception of the nature of sets.

Let us try a little harder to make sense of the power-set axiom. Bernays views sets in a 'quasi-combinatorial' way, 'in the sense of an analogy of the infinite to the finite': a set is determined by deciding, for each candidate member, whether it is to be included in the set or not [1935]. The decision is made independently for each candidate. The set is determined by the outcome of infinitely many decisions, rather than by any finitely expressible membership criterion. Maddy takes up this idea:

> Finite combinatorics tell us that there is a unique subcollection of a finite collection for every way of saying yes or no to each individual element. Carrying this notion into the infinite, subcollections are 'combinatorially' determined, one for every possible way of selecting elements, regardless of whether there is a specifiable rule for these selections. [1990, 102]

Obviously, this idea presupposes that, in some suitably idealised or metaphorical sense, it is possible to take infinitely many decisions. This sort of infinitistic thought experiment fits naturally with the transfinite iteration process involved in the iterative conception. The quasi-combinatorial view gives us a conception of an 'arbitrary' subset of a given set A. Perhaps (this is a further leap of faith, but a fairly natural one in this context) it makes sense to go from an *arbitrary* subset of A to *all* subsets of A, and to gather them into a set, $\mathcal{P}(A)$.

In a similar vein, Wang says that a set is a multitude for which we can form an intuitive 'overview', or can 'look through' or 'run through' or 'collect together' the elements. This leads him to the power-set axiom:

> For example, not only are the infinitely many integers taken as given, but we also take as given the process of selecting integers from this

> unity of all integers, and therewith all possible ways of leaving integers out in the process. So we get a new intuitive idealization (viz. the set of all sets of integers) and then one goes on. [1974, 182]

One very useful application of the quasi-combinatorial idea is to justify the impredicativity in the separation and replacement axioms. If a set A is available at stage α then so are all its subsets, so $\mathcal{P}(A)$ is available at stage $\alpha+1$. In particular, $\{x \in A \mid \phi(x)\}$ is available at stage α, for any set-theoretic formula $\phi(x)$. Many authors have been worried by the fact that $\phi(x)$ may contain reference to sets that have not yet been formed; it may even quantify over all sets. If we are forming sets in successive stages, how can we claim that $\{x \in A \mid \phi(x)\}$ is formed by stage α when it depends on sets that do not yet exist? For Hallett, this undermines the iterative conception fatally [1984, §6.1]. The quasi-combinatorial idea comes to the rescue here. Whenever we go from one stage to the next we form all sets consisting of sets generated previously. We do so not via expressions such as $\{x \in A \mid \phi(x)\}$ but by choosing elements individually. If this procedure is legitimate at all, it generates *all* the subsets, once and for all. As we proceed to later and later stages we can write down more expressions of the form $\{x \in A \mid \phi(x)\}$, as more sets become available to us, hence we can name more and more subsets of A, but those subsets all existed before we could name them; they were all formed by stage α at the latest. Hence the quasi-combinatorial idea provides an *axiom of reducibility* for set theory: any expression $\{x \in A \mid \phi(x)\}$, no matter what sets it refers to, is extensionally equivalent to a subset of A generated by choosing elements of A individually.

No doubt many philosophers and mathematicians would be sceptical of such infinitistic procedures (e.g., , [Parsons, 1977], but in the privacy of their own minds I believe they are influenced by this sort of idea. The widespread acceptance of the impredicative separation and replacement axioms, and the consensus that there are more sets than can be defined by finite conditions, are surely attributable to Bernays' quasi-combinatorial picture. I am willing to take infinitistic thought experiments seriously and discuss them without embarrassment (see §4). The snag is that the rules of the game do not seem to be very clear. If it makes sense to survey all the possible ways of making a subset of A, and to make a set out of them, why does it not also make sense to survey the entire iterative process of generating the ordinals and to make set out of them all? Indeed, Cantor himself seemed to countenance such a completed traversal of the ordinals, while simultaneously denying that the ordinals formed a set [1899a]; so perhaps this is not, after all, the basis for the distinction between sets and inconsistent multiplicities.

However this may be, it is a fundamental *assumption* of the iterative conception that it is possible to 'finish' the process of generating all the subsets of A, but not possible to 'finish' the process of generating all the stages. This suggests, even more strongly than in the limitation of size conception, that it would be appropriate to use intuitionistic logic for quantifiers that range over all sets [Pozsgay, 1971; Tharp, 1971]. Perhaps the moral of the paradoxes is that the concept of set is *indefinitely extensible*, in Dummett's sense [1994]; our attempts to grasp it always have an incomplete and provisional character. We can never say ‘for all sets’, but merely ‘for all sets we have managed to encompass so far’.

Absolute infinity, then, must be viewed as a potential infinity, in rather the same way that constructivists view countable infinity. This does not sit easily with some of Cantor’s arguments. For Cantor claimed that every potential infinity implies an actual infinity; that is, anyone who accepts a certain domain as a potential infinity is on a slippery slope that can only lead to acceptance of it as an actual infinity [Hallett, 1984, §1.2]. If we apply this argument to absolute infinity then we are forced to accept the transfinite iteration as ‘completable’, thus blurring the distinction between transfinite and absolute infinity.

2.6 Conclusions on actual infinity

Traditional philosophy attached great importance to the distinction between the finite and the infinite: the finite was the proper domain of human reason; the infinite eluded human understanding and could only be hinted at feebly by way of potential infinity. Cantor accepted this basic framework but refined it by adding an intermediate category, the transfinite (the infinite being rechristened ‘absolute infinity’). The transfinite is like the finite in being ‘actual’, a ‘unity’, and amenable to mathematical comparison and calculation, but it includes mathematical domains previously classed as infinite. Indeed, it could be said that Cantor’s aim was to shift the boundary between finite and infinite upwards so that mathematically useful systems such as $\mathbf{N}$, $\mathbf{R}$ and $\mathcal{P}(\mathbf{R})$ would lie on the finite side [Hallett, 1984, §1.3]. Bijections and similarities were the instruments by which Cantor measured, carved up, and regulated the transfinite.

There is no denying Cantor’s success in generating an appealing and coherent mathematical theory, in a topic where his predecessors had seen only confusion and paradox. But is his notion of transfinite ultimately sustainable? All ways of understanding the transfinite rest on an analogy with the finite. I have examined two in detail: a static view, in which some pluralities are seen as limited in size and others are seen as too big to have a size; and a dynamic view, in which sets are generated iteratively (I am leaving aside here the class view of sets, which has no transfinite/absolute distinction). Both views are plagued by a suspicion of semantic indeterminacy. On the iterative conception, we encounter embarrassing gaps in the iteration process: we have to accomplish the step from finite stages to stage ω and from countable stages to stage ω_1 by specific hypotheses. On the limitation of size view, how do we show that $\mathbf{N}$ and $\mathcal{P}(\mathbf{N})$ are limited rather than absolutely infinite? The problem is not that the full facts are not available to us (we know exactly what $\mathbf{N}$ and $\mathcal{P}(\mathbf{N})$ are supposed to be; no information is hidden), or that the required proofs have yet to be discovered, but that our notion of transfinite iteration or limited size does not seem substantial enough to determine an answer. Our inability to evaluate cardinal exponentials, $2^{\aleph_\alpha}$ or $\aleph_\alpha^{\aleph_\beta}$, may simply be a symptom of an underdetermination of meaning.

The problem of distinguishing between transfinite and absolute infinity is really an instance of what I termed the 'horizon problem' in §1. To give a global account of sets we need to say something about absolute infinity; our account will, almost irresistibly, lead to our treating absolute infinity as an *actual* infinity, and hence will run the risk of conflating it with the transfinite. The horizon problem is especially acute for Cantorians as they must justify converting the potential infinity $0, 1, 2, 3, \ldots$ into ω, and accepting ω as an object, while resisting a similar treatment for absolute infinity.

There are other possible approaches to the question that I have not discussed. I have ignored the theological side of Cantor's thinking on infinity [Hallett, 1984, §§1.1, 1.3–4], as I find it tends to make the issue of absolute infinity (and whether it is an actual infinity) even more murky. I have also not mentioned the approach of von Neumann [1925], in which absolutely infinite multiplicities are admitted as genuine objects but are not allowed to occur as members of any multiplicity. That is, the set V of all sets exists, but we are banned from forming the set $\{V\}$. This avoids the paradoxes successfully, but it is profoundly non-Cantorian and I am unable to see any philosophical rationale for it. Nor have I discussed the recent attempts to found set theory on plural quantification and second-order logic [Pollard, 1990], as this seems merely to rephrase the problems in a different language.

Sceptics will say that the central notion of transfiniteness is ill-defined and that the attempt to understand infinity by analogy with the finite is a basic category error. Supporters of set theory will point to the large measure of agreement about the treatment of small transfinite sets such as $\mathbf{N}$, $\mathbf{R}$ and $\mathcal{P}(\mathbf{R})$ and will maintain that any haziness about absolute infinite does not compromise the ordinary practice of mathematics, where we deal only with particular 'given' transfinite sets on any particular occasion.

3 POTENTIAL INFINITY

3.1 Introduction

The 'potential' view of infinity arises naturally from a constructivist philosophy of mathematics. By *constructivism* I mean the doctrine that constructions are the subject-matter of mathematics. For Brouwer, this means *mental* constructions, generated in the mind through the primal intuition of 'two-ity' (a process of successively forming ordered pairs of mental events); see [van Stigt, 1990]. Heyting sees constructions in a more general way, as arising from 'the possibility of an indefinite repetition of the conception of entities' [1956, p. 13]. Bishop sees natural numbers as the prototype of all constructions [1967; 1970]. Hilbert was a constructivist in the 1920s (see his [1925] and [1927]); his constructions were configurations of concrete symbols:

> as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation, certain extralogical concrete objects that are intuitively present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with these objects, as something that neither can be reduced to anything else nor requires reduction. [Hilbert, 1925]

For my purposes in this chapter, it does not matter whether the constructions are mental or physical; what matters is that they are built up out of discrete building blocks by applying certain construction operations repeatedly. The word 'construction' is not meant in a metaphorical sense, divorced from everyday usage; it includes things such as houses, which are constructions of bricks and mortar, but also ranges more widely over all mental or physical structured assemblages of elements.

When we speak of constructions we mean *tokens* rather than *types*: Hilbert considers physical inscriptions rather than their abstract shapes [1925]; Brouwer considers mental tokens (events in particular people's minds at particular times) rather than mental types (abstract thoughts that can be instantiated in people's minds). Nevertheless, we treat construction tokens of the same type as interchangeable, so it will often *sound* as if we are talking about construction types: when we speak of 4 we are referring ambiguously to any token of type "4", and when we say '4 has a unique successor' we really mean 'all successor tokens of "4" tokens are of the same type'.

The simplest kind of construction is the natural number, constructed from 0 by applying the successor operation S repeatedly. Any repetitive process can be considered to provide a system of natural number tokens: 0 is the initial state and S is the application of one further iteration of the process. The infinity of natural numbers is understood in a modal way: given any natural number (token) n, it is *possible* to construct the successor $S(n)$. Of course, at any one time only a finite supply of natural number tokens has yet been constructed, but this does not matter: in constructive mathematics we are not concerned with the actual supply of constructions (nor even the set of all possible constructions), but with the construction *process* and the fact that it can always be continued one more step. Hence we have potential infinity without actual infinity.

Constructivists generally reject actual infinity altogether. There are three possible grounds for such a rejection.

a. Ontological grounds, i.e., the claim that no actual infinities exist. This seems to have been Hilbert's position [1925; 1930; 1931].

b. Epistemological grounds, i.e., the claim that we finite beings cannot know whether actual infinity exists.

c. Semantic grounds, i.e., the claim that we finite beings cannot meaningfully refer to actual infinity. This is Dummett's view [1975b].

Notice that these grounds are mutually exclusive: if you believe that all talk of actual infinity is meaningless then you cannot even pose the question of whether actual infinity exists; if you believe that no actual infinities exist then you evidently believe the question is meaningful and answerable. Nevertheless, constructivists often try to avoid committing themselves on these fundamental philosophical matters [Heyting, 1956, chapter I; Bishop, 1967, 2; Troelstra and van Dalen, 1988, vii].

In this section I shall try to explicate the notion of potential infinity, primarily in the form of the indefinite repeatability of the successor operation on natural numbers, and to consider the objections posed by supporters of actual infinity and by feasibilists. I shall not attempt to review constructive mathematics (for which see [Bishop, 1967; Bridges and Richman, 1987; Beeson, 1985]), , but shall instead focus on the logical issues (especially the meanings of the logical constants) and the understanding of infinity.

3.2 Potential infinity versus actual infinity

It should be clear from the brief description in the previous subsection that the key idea underlying potential infinity is that of *indefinite iteration* of an operation such as successor. This idea has come under attack from supporters of actual infinity, who believe that it is ill-defined or that it presupposes actual infinity. If iteration cannot be made sense of without relying on actual infinity then the distinction between potential and actual infinity becomes blurred and the case for constructivism is seriously undermined. I shall consider these objections in this subsection.

Let us consider three possible ways of defining the concept of natural number.

1. Define **N**, the set of natural numbers, using second-order logic or set theory. There are various ways of doing this, but the general idea is to define **N** as the intersection of all sets that contain 0 and are closed under S.
2. Define **N** as the set of all things that can be produced by applying S repeatedly to 0.
3. Define the counting algorithm (the algorithm that enumerates 0, $S(0)$, $S(S(0))$, $S(S(S(0)))$, $\ldots$, without end — see the flowchart in figure 1), and interpret all statements about the concept of natural number as being really about the counting algorithm.

Options (1) and (2) involve actual infinity, as they countenance the set of all natural numbers as a completed infinity; option (3) involves potential infinity, as it avoids reference to the totality of all natural numbers. Cantor took option (2) (see §2.5), as did Dedekind in his [1872], but by [1888b] Dedekind had developed

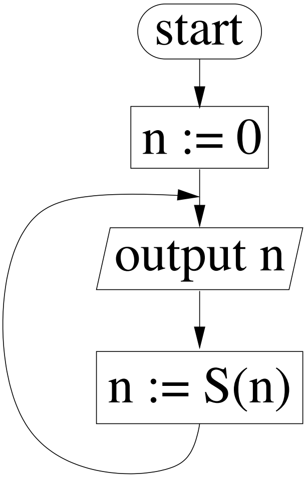


Figure 1. The counting algorithm.

his theory of 'chains' and switched to option (1). Frege also supported option (1): his theory of the ancestral relation [1879] is similar to Dedekind's theory of chains. Many modern platonists take option (1) in order to distance themselves as far as possible from iteration and potential infinity.

The choice between these options is the key point of contention between platonists and constructivists (it is much more important than the question of whether to accept the principle of excluded middle). The platonist view in favour of option (1) is supported by the following five arguments.

(i) The circularity argument. [Dedekind, 1890] claims that options (2) and (3) involve a vicious circle. In option (2) we are defining a natural number as anything that can be reached from 0 by *a finite number of iterations* of S; but it is clearly circular to define 'natural number' in terms of 'finite number'. Dedekind continues:

> The mere words "finally get there at some time", of course, will not do either; they would be of no more use than, say, the words "karam sipo tatura", which I invent at this instant without giving them any clearly defined meaning.

This argument is repeated by later platonists, such as Tait [1983] and Mayberry [2000]. The argument clearly applies against option (3) as well. Understanding the flowchart in figure 1 involves realising that one is supposed to start at the 'start' box and follow the arrows *for ever*, which presupposes infinite time and an infinite number of steps, indexed by natural numbers; hence option (3) presupposes both actual infinity and the general concept of natural number. Mayberry denounces 'operationalism' (i.e., option (3)) at length; finally he exclaims 'But what does "for ever" mean? *That* is the very point at issue!' [p385]

This argument utterly fails to impress constructivists. For them, the flowchart is a finite object (four boxes linked by four arrows). The instructions are clear: start at the 'start' box and keep following the arrows. The instructions do not contain the phrases 'finally', 'finite number' or 'for ever'. The flowchart is unusual in having a 'start' box but no 'stop' box. This absence of a 'stop' box is the true meaning of infinity; unfortunately, platonists have inflated this humble notion and made it into something mysterious. Constructivists insist that this idea of *keeping going* is immediately intelligible, does not presuppose any grasp of eternity, and is all the notion of infinity we have or need (for arithmetic). There is something about this view that seems to fit the human condition. We know how to put one foot in front of the other, even though we have no inkling of our final goal. As Bob Dylan says, the only thing we know how to do is to *keep on keepin' on*. Brouwer has attempted to explain this in terms of temporal intuition [1907], but such explanations have been generally found to be opaque and unhelpful. Perhaps the best tactic for constructivists is simply to insist that we understand algorithms, as evidenced by our ability to execute them, compare them, and express them in various programming languages, and to take this understanding as primitive.

(ii) The reducibility argument. The circularity argument usually occurs coupled with a second argument, that because natural number *can* be defined in terms of second-order logic or set theory, it *should* be so defined. Mayberry makes this claim, on the grounds that a mathematical analysis of a concept is generally preferable to a philosophical explanation of it [2000, p. 271]. Nevertheless, many twentieth-century authors have felt that to define natural number in terms of second-order logic or set theory is to misrepresent the real conceptual relations between them; as Poincaré puts it, 'This method is evidently contrary to all sane psychology' [1905]. Poincaré also accused the early logicists and formalists of circularity, as their metamathematical methods presupposed the intuition of iteration and the principle of induction, which they were claiming to prove [1905; 1906a; 1906b]. Similar reassertions of the fundamental character of iteration have been made frequently since then [Weyl, 1921; Skolem, 1922; Heyting, 1956,13]. There is no dispute about the fact that natural numbers can be *characterised* in terms of Dedekind's chains or in terms of sets, and that the principle of induction can then be proved. We can all admire the ingenuity of this trick; the point at issue is whether it is convincing as a conceptual analysis, whether this should be taken as the *definition* of number, whether it represents a *reduction* of number to something more fundamental. Skolem says:

> Set-theoreticians are usually of the opinion that the notion of integer should be defined and that the principle of mathematical induction should be proved. But it is clear that we cannot define or prove ad infinitum; sooner or later we come to something that is not further definable or provable. Our only concern, then, should be that the initial foundations be something immediately clear, natural, and not open to question. This condition is satisfied by the notion of integer

> and by inductive inferences, but it is decidedly not satisfied by set-theoretic axioms of the type of Zermelo's or anything else of that kind; if we were to accept the reduction of the former notions to the latter, the set-theoretic notions would have to be simpler than mathematical induction, and reasoning with them less open to question, but this runs entirely counter to the actual state of affairs.

Thus the controversy reduces to the question of which starting point, iteration or set theory, is the more philosophically secure and mathematically workable.

(iii) The argument from extraneous empirical circumstances. Platonists believe that counting and other algorithmic processes are unsuitable as a basis for arithmetic since they are inextricably entangled with irrelevant empirical considerations, such as who is doing the counting, what materials they are counting with, and under what circumstances [Russell, 1903, §§109, 133]. Of course, platonists accept algorithms as useful methods for discovering arithmetic facts, but not as providing a foundation for arithmetic facts. Frege, discussing the concept of 'y following in a sequence after x', rejects the idea of defining it iteratively, in terms of shifting attention step by step from x to y:

> Now this [the iterative definition] describes a way of discovering that y follows, it does not define what is meant by y's following. Whether, as our attention shifts, we reach y may depend on all sorts of subjective contributory factors, for example on the amount of time at our disposal or on the extent of our familiarity with the things concerned. Whether y follows in the ϕ-sequence after x has in general absolutely nothing to do with our attention and the circumstances under which we transfer it ... My definition [option (1)] lifts the matter onto a new plane; it is no longer a question of what is subjectively possible but of what is objectively definite. [1884, §80]

(iv) The feasibility argument. An obvious objection to options (2) and (3) is that we cannot, in fact, continue an iteration indefinitely. We shall sooner or later drop dead, or run out of paper, or encounter some other physical obstacle. For example, we cannot count up to 10^{100}; yet 10^{100} is certainly a natural number. Hence number cannot be based on counting. This argument originates with Cantor [Hallett, 1984, p. 27], and has been repeated with great frequency ever since [Bernays, 1935; Tait, 1986; George, 1988].

(v) The argument from error. Another obvious objection, though surprisingly less popular, is that any counting process is subject to error. An error is a discrepancy between the count and the true sequence of numbers. The fact that such discrepancies occur, or are even conceivable, demonstrates that number is founded on something other than counting.

I am not including any constructivist counter-arguments to the last three objections here, as they are better postponed to the subsection on feasibilism (§3.5). For the moment I shall just content myself with the bland response that constructivists base arithmetic on an 'idealised' counting process, free of empirical considerations, practical limitations and errors.

3.3 Mathematics as a process of construction

Let us suppose that we are persuaded by the constructivist side of the argument in §3.2. Like Baire [1905], we reject the idea of an infinite set, conceived as like a bag of marbles, and replace it with the idea of a construction process (i.e., option (3)). How then are we to understand statements involving infinite quantifiers? For example, consider the statement $\exists n\, OP(n)$, where the variable n ranges over the natural numbers and $OP(n)$ means that n is an odd perfect number (I shall use the word 'statement' for a formula in a mathematical language based on predicate calculus, or for its informal equivalent). There is an algorithm for testing whether a given natural number is odd and perfect, so the meaning of the predicate OP is clear. The problem is the quantifier $\exists n$, which ranges over an infinite set. A platonist would understand $\exists n\, OP(n)$ as like an infinite disjunction $OP(0) \vee OP(1) \vee OP(2) \vee \cdots$ and would say that it is true iff at least one of the disjuncts is true. This explanation is not open to us. Constructivists regard $\exists n\, OP(n)$ not as *asserting* that there is an odd perfect number, but as expressing the *intention* to find an odd perfect number or the *problem* of finding an odd perfect number [Heyting, 1931; Kolmogorov, 1932]. If N is an odd perfect number then we may say 'N achieves the intention $\exists n\, OP(n)$' or 'N solves the problem $\exists n\, OP(n)$', or 'N is a proof of $\exists n\, OP(n)$'; these are three ways of expressing the same judgement, which I shall also write using the notation

$$N \vdash \exists n\, OP(n).$$

(This is a special sense of the word 'proof' and the symbol '$\vdash$', peculiar to constructivism and separate from the way they are used in platonistic mathematics.) The meaning of $\exists n\, OP(n)$ is given by specifying the conditions under which a given construction N counts as a 'proof' of it. It is easy to specify these conditions in this example: even though we do not know any odd perfect number, we certainly do know how to test whether a given number is odd and perfect.

Essentially the same idea is expressed by Weyl [1921] and Hilbert [1925]. Weyl calls an existential statement a 'judgement abstract', and Hilbert calls it a 'partial proposition'. They regard $\exists n\, OP(n)$ as an incomplete judgement (one could write it as $OP(-)$), which needs to be supplemented by a construction, N, to make a complete judgement, $OP(N)$. We can say that a 'proof' of a statement is a construction that, when inserted into the statement, makes a true judgement.

Mathematics, then, is simply an activity of finding constructions that prove statements. A typical mathematical judgement is either of the form 'P is a proof of A' (i.e., $P \vdash A$) or 'given any construction x of a certain kind, $f(x)$ is a proof of $A(x)$'. Note that the latter judgement involves a kind of universal quantifier: one is tempted to write it as $\forall x\ f(x) \vdash A(x)$. In fact, even the simple judgement $P \vdash A$

involves an element of generality: it means that if any mathematician repeats the construction P, at any time, without error, then they must reach the same conclusion, that it proves A. Hence a constructive mathematical judgement is not merely an empirical report of a particular construction episode (as sometimes suggested by Brouwer, Placek [1999] and Heyting [1956, pp. 3, 8]), but involves an aspect of generality, timelessness, objectivity and necessity.

How is this generality to be understood? Weyl and Hilbert say that we can understand a simple universal statement 'for any n, $A(n)$' (where n ranges over $\mathbf{N}$) as a kind of schema, that *becomes* a judgement $A(N)$ when a numeral N is substituted into it. We can interpret this in terms of Russell's distinction between 'for all' and 'for any' (1908, II). 'For all n' is the universal quantifier $\forall n$ of predicate calculus; $\forall n\ A(n)$ is understood by platonists as an infinite conjunction $A(0) \wedge A(1) \wedge A(2) \wedge \cdots$. 'For any n, $A(n)$' is a singular schematic judgement, meaning that $A(n)$ holds *regardless* of the value of n. It seems best to say, therefore, not '$\forall x\ f(x) \vdash A(x)$', but 'for any x, $f(x) \vdash A(x)$' (this is consistent with the characterisation of intuitionistic argument given by Herbrand [1931, footnote 3]).

To summarise, we hope to interpret constructivist mathematics in terms of two ideas:

- that a mathematical statement, A, is an incomplete judgement, something needing to be supplemented by a construction, P, to form a complete judgement, $P \vdash A$;
- that generality in mathematical judgements is to be understood in terms of 'for any' rather than 'for all'.

3.4 The meaning of the logical constants

So far we have only considered a very simple example of a mathematical statement, $\exists n\ OP(n)$. What happens when other logical constants are used, and when the logical constants occur in combination? Hilbert's approach was to interpret each mathematical theory as a whole: he regarded formal mathematical theories as devices for deriving true decidable statements [1925; 1927]. Other constructivists, however, have sought to attach a meaning to individual statements, in line with the idea that the meaning of a statement is given by specifying what type of construction counts as a proof of it. Let us confine our attention to formulae in the language of Peano arithmetic, for the sake of simplicity. The atomic formulae are decidable: they can be evaluated by a computation to produce the result *true* or $false$. Hence, for an atomic formula, we may consider a trivial construction, such as 0, as a proof if the formula evaluates to *true*, and no construction as a proof if the formula evaluates to $false$.

The meaning of the logical constants is determined as follows.

a. To prove $A \wedge B$ one must prove A and prove B.

b. To prove $A \vee B$ one must select one of the two disjuncts and prove it.

c. To prove $A \Rightarrow B$ one must show how, given a proof of A, to obtain a proof of B.

d. To prove $\exists n\ A(n)$ one must select a natural number N and prove $A(N)$.

e. To prove $\forall n\ A(n)$ one must show how, given a natural number N, to prove $A(N)$.

(I have not included a clause for $\neg A$ because it is convenient to define $\neg A$ as $A \Rightarrow \mathit{false}$.) I venture to call this the *intended meaning* of the logical constants, since so many constructivists have given explanations of the logical constants approximately along these lines indexHeyting, A.[Kolmogorov, 1932; Heyting, 1956, §§7.1.1, 7.2.1; Bishop, 1967, §1.3; van Dalen, 1973, §2.1; Dummett, 1977, §1.2; Beeson, 1985, §II.6; Bridges and Richman, 1987, §1.3; Troelstra and van Dalen, 1988, chapter 1, §3.1]. Indeed, if this is not the intended meaning, it is hard to imagine what else could be intended. However, it is in the attempt to render this more precise that differences between constructivists emerge.

In the first place, the above account is phrased as a definition of a verb ('prove'), but in clause (c) it slips into use of the noun ('proof'). In the previous subsection we characterised a mathematical statement as an incomplete judgement, and a proof as the construction that has to be inserted in the gap to make a true judgement. In the light of this it is natural to take 'proof' rather than 'prove' as the definiendum. The above clauses need to be rewritten accordingly.

However, before doing so we need to decide what *type* of thing a proof is. In ordinary speech the word 'proof' denotes a sequence of judgements, proceeding from axioms or hypotheses to a conclusion. The formal counterpart of this is a tree of formulae, having axioms or hypotheses at the leaves and the conclusion at the root. I shall refer to these as *informal* and *formal derivations*, respectively. If a proof is a formal derivation then it is natural to rewrite clauses (a)–(e) as follows.

a′. A proof of $A \wedge B$ consists of a proof of A, a proof of B, and a final inference $\frac{A \quad B}{A \wedge B}$.

b′. A proof of $A \vee B$ consists of either a proof of A followed by a final inference $\frac{A}{A \vee B}$, or a proof of B followed by a final inference $\frac{B}{A \vee B}$.

c′. A proof of $A \Rightarrow B$ consists of a proof of B from the hypothesis A.

d′. A proof of $\exists n\ A(n)$ consists of a proof of an instance $A(N)$ followed by a final inference $\frac{A(N)}{\exists n\, A(n)}$.

e′. A proof of $\forall n\ A(n)$ consists of a proof of $A(n)$ followed by a final inference $\frac{A(n)}{\forall n\, A(n)}$.

It appears from this that constructive proofs are formal derivations in a natural-deduction logic, using only introduction rules. But this is too restrictive: we must allow some use of elimination rules as well if formulae such as $(A \wedge B) \Rightarrow (B \wedge A)$ are to be provable. So perhaps it is better to take proofs as normal natural-deduction derivations, as suggested by Dummett [1977, p. 396].

However, there is surely something circular about defining a proof as a formal derivation when it is our intention to use the proof clauses (a)–(e) to *justify* the axioms and rules of inference used in formal derivations. Moreover, both formal and informal derivations seem very different from the special notion of proof introduced in §3.3, in which a proof is simply a construction that fills a hole in an incomplete judgement, and may be something as simple as a natural number. Dummett acknowledges this difference, referring to a proof in the sense of clauses (a)–(e) as a *canonical proof* and a mathematical derivation (whether formal or informal) as a *demonstration*; a demonstration provides the means for constructing a canonical proof [1977, 391-4].

Accordingly, many authors would write the explanations of the logical constants in the following form.

a$''$. $P \vdash A \wedge B$ iff $P = (Q, R)$, where $Q \vdash A$ and $R \vdash B$.

b$''$. $P \vdash A \vee B$ iff $P = (i, Q)$, where $i = 0$ and $Q \vdash A$, or $i = 1$ and $Q \vdash B$.

c$''$. $P \vdash A \Rightarrow B$ iff P is a function such that, for any Q, if $Q \vdash A$ then $P(Q) \vdash B$.

d$''$. $P \vdash \exists n\ A(n)$ iff $P = (N, Q)$, where $Q \vdash A(N)$.

e$''$. $P \vdash \forall n\ A(n)$ iff P is a function such that, for any natural number N, $P(N) \vdash A(N)$.

There are many theories of constructive proof along the lines of clauses (a$''$)–(e$''$). I shall not attempt to describe them, but shall rather review the common dilemmas they face.

Theories of constructive proof may be classified into *bottom-up* theories, which start with well-understood mathematical constructions such as natural numbers and try to use them to define a concept of proof based on clauses (a$''$)–(e$''$); and *top-down* theories, which take the full pre-theoretic concept of constructive proof and try to analyse it to make it more explicit and intelligible.

The simplest bottom-up theory is Kleene's concept of realisability [1945]. The proofs are natural numbers, the functions in clauses (c$''$) and (e$''$) are partial recursive functions coded as natural numbers, and the pairs are also coded as natural numbers. Gödel's [1958] 'Dialectica' interpretation, Scott's [1970] theory, and Martin-Löf's [1984] intuitionistic type theory are also bottom-up theories, though involving higher-type constructions built upon the natural numbers.

A feature of all these systems is that the $\vdash$ relation is undecidable. This should be clear from clauses (c$''$) and (e$''$), since it is undecidable whether a function satisfies the condition to be a proof of $A \Rightarrow B$ or $\forall n\ A(n)$. As Beeson points out, this deviates considerably from the usual notion of constructive proof [1985, 281-2]. Consider for example Goldbach's conjecture, $\forall n\ G(n)$, where $G(n)$ is the decidable statement that $2n + 4$ is the sum of two prime numbers. If Goldbach's conjecture is constructively valid then we have

$$f \vdash \forall n\, G(n),$$

where f is a function mapping any natural number n to 0. (Recall that a decidable statement such as $G(n)$ has 0 as a proof iff it is true.) But no one would accept f as a 'proof' of Goldbach's conjecture. The problem of determining whether $f \vdash \forall n\, G(n)$ is as hard as the problem of solving Goldbach's conjecture; f contributes nothing to the solution of the conjecture. The statement $\forall n\, G(n)$ is an incomplete judgement, but f fails to fill the gap; it leaves the gap as wide as ever. Hence it is often proposed that a proof ought to include enough information to make it clear that it *is* a proof; in other words, the $\vdash$ relation should be decidable [Kleene, 1945, §2; Kreisel, 1962, §4; Dummett, 1977, p. 12]. (See [Sundholm, 1983] for a contrary view, however.)

There are some bottom-up theories in which $\vdash$ is decidable, such as the early version of Martin-Löf's type theory [1975]: functions are given by terms in a special language, and by reducing the term to a normal form it can be seen whether it is a proof of a given formula. Indeed, the view that proofs are formal derivations (clauses (a$'$)–(e$'$) above) may also be regarded as a theory of this type. Decidability, however, is bought at the price of a limited language for defining functions. By diagonalising out of the class of expressible functions we can define a function, not expressible in the language, that transforms any proof of a formula A to a proof of a formula B. In such a case, we are able to derive proofs of B from proofs of A, yet we do not have a proof of $A \Rightarrow B$. This defeats the whole purpose of constructive implication: the idea of $A \Rightarrow B$ was to express the problem of converting any given proof of A to a proof of B (clause (c)).

Hence all bottom-up theories have difficulties conforming to the intended meaning of proof. Top-down theories proceed in the other direction, beginning with clauses (a)–(e) or (a$''$)–(e$''$) and trying to infer the nature and structure of proofs. These theories distinguish between 'concrete' mathematical constructions (involving natural numbers, recursive functions, and similar discrete things) and 'abstract' mathematical constructions (involving mentalistic notions of proof, meaning and rules). (Beeson [1985, §§III.8, VI.10] uses the word 'rule' to mean a procedure whose application requires understanding of meaning, but not creativity or free will.) It is not assumed that abstract constructions can be reduced to concrete ones; in particular, it is not assumed that all rules mapping one natural number to another are recursive functions. It is assumed that $\vdash$ is decidable, in the sense that a human can decide whether $P \vdash A$ by attending to the meaning of P and A; but it is not assumed that $\vdash$ is mechanically decidable, i.e., recursive.

In order to make $\vdash$ decidable, it is necessary to add 'supplementary data' or 'second clauses' into the clauses (c$''$) and (e$''$):

c$'''$. $P \vdash A \Rightarrow B$ iff $P = (E, F)$, where F is a function and E is evidence that, for any Q, if $Q \vdash A$ then $F(Q) \vdash B$.

e$'''$. $P \vdash \forall n\ A(n)$ iff $P = (E, F)$, where F is a function and E is evidence that, for any natural number N, $F(N) \vdash A(N)$.

Thus every proof carries its own evidence that it really is a proof. But what is evidence, exactly? Beeson [1985, p. 39] and Díez [2000] believe that evidence is simply proof. Thus the definition of $\vdash$ is circular, rather like Tarski's platonistic definition of truth of a formula in a model. Beeson finds this circularity untroubling, since constructive proof is a fundamental concept and needs no explaining. Nevertheless, this leaves us with a rather obscure notion of proof, with the proof relation characterised in terms of an ill-delimited totality of all proofs (clause (c$'''$)). As Beeson admits [402], this disrupts the attempt to define $P \vdash A$ by structural induction on the formula A.

Most authors prefer to construe evidence as something simpler and more fundamental than proof. Note that the 'for any' quantifier used in clauses (c$'''$) and (e$'''$) may be understood in the sense explained in §3.3, rather than as $\forall$. All we need to assume is that it is decidable whether any given E is evidence for a judgement of the form 'for any x, $C(x)$', where $C(x)$ is a decidable statement or a conditional involving decidable statements; then $\vdash$ becomes decidable. The definition of $\vdash$ becomes non-circular, since we are defining proof of a formula in terms of proof of its syntactic components, using a predefined concept of evidence. All we need to complete the story is a philosophical account of evidence.

Theories of this sort are provided by Kreisel [1962; 1965] and Goodman (1970, 1973). Unfortunately, neither tells us anything about what evidence is. Worse, it turns out that the assumption that evidence is decidable and sound leads to a self-referential paradox [Goodman, 1970, §9]. Goodman attributes this to a vagueness in the description of how constructions are built up [§10], and responds by stratifying the universe of constructions into levels. The bottom level consists of constructive rules, the next level consists of rules and proofs about rules, and so on. Goodman restricts the scope of each item of evidence to apply to constructions of a particular level. He then finds that he needs to posit a reducibility operator (§11), acting rather like the reducibility axiom in Russell's ramified type theory [1908], and suffering of course from the same implausibility.

This unexpected need for stratification and reducibility is related to a perceived impredicativity in the definition of constructive implication. If one believes that proofs are composed of formulae (as they would be if they were natural-deduction derivations), then to understand $A \Rightarrow B$ one is required to survey all proofs of A, some of which may contain $A \Rightarrow B$. This is clearly an impredicativity. To avoid this, Dummett proposes that, for any statement A, there should be an upper bound on the complexity of the proofs of A that need be considered, a bound that increases with the syntactic complexity of A [1977, §7.2, especially 394-6]. The impredicativity does not arise if proofs are not composed of formulae, but the complexity bound is still desirable as it makes the task of surveying all proofs of A more tractable. Goodman also makes the proposal of a complexity bound; and he points out that if one applies it to the case where A is itself an implication $C \Rightarrow D$, then it implies that all functions mapping proofs of C to proofs of D may be assumed to be of bounded complexity [1970, §11]. This itself is a sort of axiom of reducibility.

The foregoing argument suggests that stratification and reducibility are necessary to the coherence of constructive implication; unfortunately it does nothing to make them plausible.

It should be clear now that constructivism is in serious difficulties in its attempt to make sense of its fundamental concept of proof. Bottom-up theories, appropriate to constructivists such as Markov [1968] and Bishop [1970], fail to do justice to the intended meaning of proof (i.e., clauses (a)–(e)). Top-down theories, which are more appropriate to intuitionism, rely on our supposed ability to grasp and quantify over an ill-defined universe of proofs. They appeal to *ad hoc* assumptions of stratification and reducibility to make constructive implication manageable. They make heavy use of mentalistic notions such as 'understanding of meaning' and 'humanly computable rule', which seem to entangle them with intractable problems in the philosophy of mind.

Díez [2000] offers a hybrid solution that uses a concept of realisability similar to Kleene's, but also uses 'second clauses' at the outermost level:

$$P \vdash A \text{ iff } P = (E, N), \text{ where } E \vdash \text{`}N \text{ realises } A\text{'}.$$

This gives rise to an odd notion of implication: there will be cases in which we know how to transform a proof (E, N) of A to a proof (E', N') of B, yet N' may depend on both E and N, hence there is no function transforming N to N'. In this case, $A \Rightarrow B$ is not realised and hence is not proved. Thus we have a violation of clause (c): we know how to convert a proof of A into a proof of B, yet we have not proved $A \Rightarrow B$. As with the decidable bottom-up theories, this undermines the very purpose of implication. Díez is well aware of this phenomenon but does not seem to see it as a problem. In addition, Díez's proposal suffers from the same risk of circularity as the top-down theories, but in an especially clear form, since 'N realises A' may be of arbitrary logical complexity.

My own theory of proof can be found in Fletcher [1998]; I shall not attempt to describe it here, except to say that its aim is to combine the explicitness of the bottom-up approach with full adherence to the intended meaning (clauses (a)–(e)), using 'second clauses' and a theory of evidence.

3.5 Feasibility

It is time to return to a problem left over from §3.2. A constructivist believes that the subject matter of arithmetic is iterative processes in general, of which counting is a typical example. In §3.2 I stated an obvious objection to this view: that all iterative processes are subject to practical limitations; it is not feasible to count up to 10^{100}. This objection is used by platonists to expose the incoherence of constructivism, but it can also be made by those who genuinely believe that arithmetic should be limited to feasible numbers, numbers one can actually count up to. Such people are known as strict finitists, ultra-finitists, ultra-intuitionists or feasibilists Yessenin-Volpin, 1970; Wright, 1982].

The feasibilists accuse the constructivists of making an unwarranted and ill-defined idealisation from small finite examples to arbitrary finite examples. This accusation is rather reminiscent of the charge brought by the constructivists against the platonists, that of making an unwarranted and ill-defined extrapolation from the finite case to actual infinity. Some authors have claimed that there is a very close analogy between the feasible-to-finite idealisation and the finite-to-infinite extrapolation [Wright, 1982; George, 1988]. If so then constructivism is in trouble.

Strangely, constructivists have given little attention to this serious objection. Troelstra and van Dalen admit that it is a genuine problem [1988, p. 851]. Dummett has argued that feasibilism is incoherent, because the predicate 'feasible' is vague and hence prone to the ancient Greek paradox of the heap [1975a]. Constructivism, of course, is not similarly dependent on any vague predicate. This certainly disrupts the analogy that Wright and George are attempting to draw, but it does nothing to justify the constructivist's idealisation.

Let me try to provide the missing justification. Imagine that we are setting out to develop a theory of some counting process, such as the writing of a sequence of strokes $||||||||\cdots$; we intend to take account of all the practical factors that might affect this process, such as running out of paper or being struck by a meteorite. We first note that our theory will contain both necessary and contingent propositions. Examples of necessary propositions are:

> 'To write ten strokes entails writing five strokes, twice.'
> 'To write two strokes, followed by two more, is to write four strokes.'

Examples of contingent propositions are:

> 'It is practically feasible to write 500 strokes.'
> 'The likelihood of a meteorite strike while one is writing 500 strokes is very low.'
> 'If one attempts to copy a sequence of 500 strokes one is very likely to write too many or too few by mistake.'

Note that the contingent propositions involve vague predicates and miscellaneous empirical considerations, whereas the necessary ones do not. Note further that a necessary proposition is never logically dependent on contingent propositions. Hence the necessary propositions form a self-contained logical system: their logical inter-relationships can be studied without reference to the contingent propositions. The necessary propositions all have a conditional character: they do not contain any outright assertion that any sequence of strokes can be written (after all, even writing a single stroke might be practically very difficult in some circumstances).

Our theory of stroke-writing, therefore, falls apart neatly into two parts: a self-contained necessary theory, and a contingent theory that surrounds it and links it to our general knowledge about the physical world. The necessary theory is, of course, arithmetic. An important advantage of this decomposition is that the necessary theory can be transferred to other iterative situations. For example, if we wanted a theory of, say, building towers of cardboard boxes we could use the same necessary theory and couple it to a new contingent theory. In short, arithmetic is a theory of the necessary aspects of all iterative processes. The application of arithmetic to a particular iterative situation is a two-step process: first develop the arithmetic facts, and then take account of contingent factors. From this point of view the proposal to insert feasibility considerations into arithmetic is completely misconceived.

This account answers two objections left over from §3.2: the argument from extraneous empirical circumstances and the feasibility argument. It can also be extended to answer the remaining objection, the argument from error. The fact that we are able to repeat a calculation, compare the two versions, identify discrepancies, and determine which version is erroneous, gives an operational meaning to the concept of error. It gives us a grip on the question of objectivity, the distinction between a calculation's being correct and its merely seeming to be correct.

The feasibility issue is so important that I think it worthwhile to make the same point again by an analogy from a different field. In theoretical linguistics, a distinction is drawn between competence and performance [Chomsky, 1972, chapter 5]. The *competence* of a speaker of a language is their knowledge of its vocabulary and grammatical constructions. The *performance* of a speaker is their actual linguistic behaviour, the sentences they are able to produce or understand under various circumstances, and the errors they make. A highly convoluted sentence, containing many nested subordinate clauses, might be within a speaker's linguistic competence, in the sense that they know all the words and grammatical constructions used in it, but beyond their linguistic performance, in the sense that it is too complicated for them to take in. Modelling a speaker's language abilities is a two-step process. First we model their competence, perhaps by means of a formal grammar; the grammar will usually include recursive production rules, which can be applied any number of times in a single sentence, and hence the grammar generates an infinite set of sentences. The second step is to model the speaker's performance, which is the result of an interaction between their infinite competence and finite resource constraints of short-term memory and attention.

It may seem paradoxical that, in order to do justice to a speaker's real linguistic behaviour, we have to begin by constructing an idealised competence theory in which we pretend that the speaker can handle sentences of arbitrary complexity. Indeed, many theorists have rejected this two-step analysis and have attempted to model performance directly. The best discussion of the controversy I have seen is by Fodor and Pylyshyn, 1988, especially §3.1]. They point out the explanatory value of the two-step analysis: 'a speaker/hearer's performance can often be improved by relaxing time constraints, increasing motivation, or supplying pencil and paper. It seems very natural to treat such manipulations as affecting the transient state of the speaker's memory and attention rather than what he knows about — or how he represents — his language.'

These considerations generalise to other types of cognitive function, particularly arithmetic. A person's performance in an arithmetic test can be accounted for as an interaction between their arithmetic competence (the algorithms they know) and the incidental circumstances (calculating aids, distractions, frame of mind, etc.). Someone who has been taught to calculate square roots in their head can calculate *more* square roots when given pen and paper — even though they have not been taught any more mathematics. Hence it is natural to infer that their knowledge of extracting square roots exceeds what they are actually able to demonstrate with particular materials.

Nevertheless, in arithmetic as in linguistics, there are those who maintain that we should develop a theory of performance directly, without going via competence; this view is feasibilism.

Let us attempt to develop a feasible theory of numbers, based on the principle that a number is feasible iff we can write that many strokes. (The word 'we' may be construed as referring to a particular person or the whole mathematical community at a certain time.) The first difficulty is that the obstacles that limit our stroke-writing are so various and variable. If our pencil point breaks then we are suddenly prevented from proceeding; there is a number n such that n is feasible but $S(n)$ is infeasible. Other obstacles, such as fatigue, boredom, or a shrinking piece of chalk, operate in a more gradual way. Sazonov [1995, §1] considers a theory of the former type, in which there is a largest feasible number; he claims that 'recursion theory relativized to such a finite row of natural numbers is essentially the theory of polynomial-time computability'. Most theories of feasibility, however, are of the latter type, satisfying the axiom

$$F(n) \Rightarrow F(S(n)),$$

where $F(x)$ means that x is feasible [Yessenin-Volpin, 1970; Parikh, 1971]. In this case the feasible numbers should satisfy Peano's axioms. One may view F as a vague predicate if one wishes. A popular way of obtaining a theory of feasibility is by adapting some system of nonstandard arithmetic. We can set up a theory of nonstandard arithmetic very simply by introducing into first-order Peano arithmetic a constant 'c', with a list of axioms '$N < c$' for every numeral N. Such a theory is consistent, since every finite subset of the axioms is consistent, so it has a model, containing an infinite natural number c as well as the standard numbers $0, 1, 2, \ldots$. Now, if we alter the terminology, substituting the word 'feasible' for 'standard', then we can re-construe this as a theory of all the natural numbers, containing the feasible numbers as a subsystem. The full natural numbers satisfy the principle of induction

$$(A(0) \wedge \forall n\, (A(n) \Rightarrow A(S(n)))) \Rightarrow \forall n\ A(n)$$

for any formula $A(x)$ not containing the feasibility predicate F. If we extend the original theory to include sets, the nonstandard theory will have *standard*, *internal* and *external* sets (in the usual sense of nonstandard set theory). When we transfer this to a theory of feasibility, the internal sets are simply called ‘sets’ and the external sets are called ‘classes’. For example, there is a (proper) class of all feasible numbers, which is a subclass of the set of all numbers less than c.

The most comprehensive attempt at providing a foundation for mathematics along these lines is *alternative set theory* (AST), developed by Vopěnka [1979] and Sochor [1984]. For readers unfamiliar with AST, a helpful ultrapower model of AST in conventional set theory is provided by Pudlák and Sochor [1984]. It turns out that in AST the feasibility predicate can be defined in terms of classes: a natural number n is feasible iff all subclasses of $\{0, 1, \ldots n\}$ are sets. AST provides a useful new view of the relation between the continuum and discrete structures, in terms of infinitesimals and indiscernibility relations (in fact, this can be done in any theory containing two models of arithmetic, one nested inside the other).

Feasibility theories of this type may be considered unsatisfying, as they say nothing about the size of infeasible numbers such as c; indeed, they allow us to prove that numbers such as 10^{1000} are feasible (by applying $F(n) \Rightarrow F(S(n))$ repeatedly 10^{1000} times). A bolder type of theory inserts a specific axiom $\neg F(\theta)$, where θ is a particular closed primitive-recursive term such as ‘10^{1000}’. This of course renders the theory inconsistent. The usual way round this, introduced by Parikh [1971], is to impose feasibility restrictions in the meta-theory, i.e., to limit the derivations to those with a feasible number of steps. It is also necessary to limit the syntactic complexity of the formulae in the derivations. For example, in AST the induction axiom is limited to formulae $A(x)$ of feasible length; otherwise one could apply induction with the formula $x = 0 \vee x = 1 \vee x = 2 \vee \cdots$ (a disjunction over all feasible numbers) to show that all numbers were feasible.

A theory of this sort will assume that the feasible numbers are closed under some arithmetic operations but not others: for example, it may contain as axioms $F(m) \wedge F(n) \Rightarrow F(m + n)$ and $F(m) \wedge F(n) \Rightarrow F(m \cdot n)$, but not $F(m) \wedge F(n) \Rightarrow F(m^n)$. Parikh shows (§2) that if a suitable upper bound is imposed on the complexity of derivations then an axiom $\neg F(\theta)$ can be added without enabling any new contradictions to be derived. Hence the theory is consistent from a feasibilist point of view.

The snag is that the value of θ has to be very large in comparison with the bound on the complexity of derivations. For example, a fairly short derivation will give:

$$F(10), \text{ hence } F(10^2), \text{ hence } F(10^4), \text{ hence } F(10^8), \ \ldots \text{ hence } F(10^{1024}).$$

Thus the concept of feasibility embodied in the theory far outruns the concept of feasibility being applied in the meta-theory. This presumably makes the theory implausible to a feasibilist.

It is possible to bring the feasibility restrictions in the theory and meta-theory into harmony by restricting the theory rather drastically in other ways [Sazonov, 1995, §§3–4]. In fact, one has to reject the law of modus ponens and accept only normal natural-deduction derivations; and one has to reject the doubling function as infeasible.

The underlying trouble here can be stated quite simply without meta-mathematics. One would like to include the following principles in a theory of feasible arithmetic.

(i) If n is a feasible number and f is a feasible function then it is feasible to apply f n times.

(ii) The doubling function is feasible.

(iii) 1000 is feasible.

(iv) 2^{1000} is infeasible.

But of course these are inconsistent. One could make a case for rejecting any of them; but giving up (i) makes the feasible number system less self-contained and the concept of feasibility less coherent; giving up (ii) makes feasible arithmetic seem very impoverished; giving up (iii) or (iv) is contrary to our informal ideas of what is feasible.

In the end we may have to accept that feasibility is a context-dependent notion (it is feasible to write 1000 strokes, but not if each stroke is required to be twice the length of the previous one). Hence the idea of axiomatising feasibility with a single unary F predicate may be too simple-minded and may only capture some aspects of real-life feasibility constraints. Perhaps it would be fair to say that feasible arithmetic does not remove idealisation from arithmetic (as the feasibilists intended) but simply shifts the boundary of the idealisation so that the abstract notion of a bound to counting is incorporated in arithmetic.

It is significant that these formal theories of feasible arithmetic have difficulty quantifying over the feasible domain. Either the quantifiers range over all natural numbers (feasible and infeasible), or they are bounded by arithmetic terms (see §2 and §4 of Parikh [1971], respectively). This is reminiscent of the dilemma in Cantorian set theory, where one either finds oneself referring to proper classes (which officially one does not believe in) or restricting the range of each quantifier to a set (with consequent difficulties in making general statements about sets). This is simply a re-emergence of what in §1 I termed the *horizon problem*. A feasibilist needs a theory of infeasible quantities, just as a finitist needs a theory of the infinite and a Cantorian set theorist needs a theory of proper classes. The rejection of idealisation beyond actual human capabilities does not do away with the problem of infinity; it merely shifts the frontier.

3.6 Conclusions on potential infinity

The view of infinity developed in §3.2 and the concept of proof developed in §§3.3–4 require a reform of the whole structure of mathematics. Kronecker was the first to see that this required rejection of the logical principle of excluded middle, $A \vee \neg A$, in cases where A is undecidable. It also requires a modified concept of real number, based on Cauchy sequences given either by constructive rules [Bishop, 1967] or by free choice sequences [Troelstra, 1977]; the former option leads to results consistent with 'classical' platonistic analysis, but the latter option leads to powerful new theorems having no platonistic counterpart. The more general consequences for mathematics, and set theory in particular, are just beginning to be elaborated; Beeson [1985] provides a good summary. Brouwer deserves the credit for almost singlehandedly establishing the philosophical basis of constructive mathematics and carrying out the required reconstruction of arithmetic and analysis [van Stigt, 1990].

Since Brouwer, constructivism has split into many schools [Bridges and Richman, 1987]. A large part of constructive mathematics, however, is common to all schools and is developed by Bishop [1967]. Moreover, proof-theoretic investigations have shown that almost all scientifically applicable mathematics (e.g., the theory of separable Hilbert spaces) can be formalised in theories that are proof-theoretically reducible to constructive number theory [Feferman, 1998]. It might be argued, therefore, that the reasons that led to the acceptance of actual infinity at the end of the nineteenth century [see §2.1] have now been superseded. Potential infinity works just as well.

The feasibility objection to potential infinity, which seems at first sight so damaging, turns out to be untroubling. Feasible arithmetic can coexist peacefully with constructive arithmetic: the two are simply based on different idealisations.

On the other hand, we have seen in §3.4 the serious obscurities surrounding the fundamental concept of constructive proof. The meaning of any formula of the form $A \Rightarrow B$ depends on a quantification over a murky and ill-delimited totality of proofs of A. By comparison, the universe of transfinite sets seems clear and definite, and our worries about it in §2 seem to pale into insignificance. The motivation for constructivism was to avoid reliance on ungraspable infinite totalities, but we seem to have ended up worse off than we were under Cantor's regime. The difficulty of reasoning constructively about the totality of constructive proofs of a formula is the constructivist's equivalent of the horizon problem.

The more perceptive critics of constructivism have pointed out this objection [Bernays, 1935, p. 266; Gödel, 1933; 1938; 1941]. Constructivists themselves have emphasised the importance and difficulty of the problem [Dummett, 1977, §7.2; Weinstein, 1983].

Clearly, the clarification of these issues is the most pressing problem facing constructivism today.

4 PHYSICAL INFINITY

4.1 Introduction

The topic for this section is the nature of infinity as it occurs in applied mathematics and physics, in both countable form (an infinite number of particles) and uncountable form (the continua of space and time and other real-valued physical quantities such as mass and charge).

We are also concerned with the relation between mathematical infinity and physical infinity. There are two obvious ways in which they could be related. One could start with mathematical infinity and use it as a model for describing physical reality. Thus one would provide a foundation for the system of real numbers on pure mathematical grounds and then hypothesise that the structure of space and time is isomorphic to a four-dimensional manifold. The relation between the pure mathematical theory and the physical application is often explained in terms of 'bridge principles' [Peressini, 1999].

Alternatively, one could start with physical infinity and use it as a foundation for mathematical infinity. An example of this is Hellman's modal structuralism. Hellman [1989] interprets propositions of number theory as modal statements about all ω-sequences (an ω-sequence is a model of second-order Peano arithmetic, and may consist of elements of any kind, including physical objects). A proposition of number theory is defined to be true iff it is necessarily the case that it holds in all ω-sequences. Using this interpretation, Hellman [1998] has argued that if one accepts even the possibility of a physical ω-sequence then one is committed to the coherence of actual infinity in pure mathematics.

It is more common, however, to argue by analogy or thought experiment from the meaningfulness and possibility of physical infinity to the meaningfulness and possibility of mathematical infinity. To platonists it seems obvious that there could be an infinity of stars: 'It is fatuous to suppose that we know *a priori* that the stars in the heavens cannot possibly go on and on forever but that at some point in space they must come to an end' [Benardete, 1964, 31]. Thus the idea of actual infinity is surely coherent. Hence the idea of an actual infinity of mathematical objects also makes sense. This seems to undermine fatally the *semantic* objection to actual infinity [see §3.1]. Furthermore, this line of thought can be used to defend the principle of excluded middle in pure mathematics. Wittgenstein and the intuitionists have questioned whether it is justified to say that the sequence 7777 must either occur or not occur in the decimal expansion of π. Benardete [1964, 129-130] imagines an infinite row of stars, colour-coded according to the digits of π (white represents 0, black represents 1, etc., with scarlet representing 7). He asserts *a priori* that either a sequence of four consecutive scarlet stars exists or it doesn't; the conclusion then transfers to π.

A second common platonist argument is that their truth-conditional view of infinite quantifiers [see §2.1] makes sense because one could, in principle, evaluate a quantified statement by examining each instance separately. Imagine, for example, testing Goldbach's conjecture (the statement that every even number greater than 2 is the sum of two prime numbers) by testing the even numbers $4, 6, 8, \ldots$ in succession. If one proceeded at ever-doubling speed one could complete all the tests in a finite time and so give a definite truth-value to the conjecture. This sort of infinitistic thought experiment is nowadays known as a *supertask*. Weyl questioned whether such a supertask was possible and related it to Zeno's paradox of Achilles and the tortoise [1949, 42].

In a similar vein, Russell believed that our inability to run through all the digits of π is not '*logically* impossible' but merely '*medically* impossible' [1935–6]. This distinction indicates that we should disregard contingent limitations in this context as we are only concerned with conceptual coherence. It would not be appropriate to rule out the supertask by citing the limited human lifespan, the impossibility of travelling faster than light, or the atomic structure of matter, as these contingent factors do not undermine the conceptual coherence of the example and hence do not affect its validity as a thought experiment to elucidate the meaning of infinity.

Let us therefore make the following idealising assumptions:

(i) an actual infinity of bodies is possible, occupying arbitrary positions (provided they do not overlap);

(ii) any physical quantity (e.g., duration, velocity, mass, density) can take on arbitrarily large finite values or values arbitrarily close to zero;

(iii) bodies are perfectly rigid and have well-defined boundaries (this assumption can be relaxed in many cases).

This idealised world-picture can be used to probe the structure of the spatial continuum. The two key characteristics of the continuum are: *divisibility* (any line segment can be divided into two line segments), and *homogeneity* (all line segments are qualitatively alike, differing only in length). By enquiring about the ultimate parts of a line segment we can identify a number of options for the structure of the continuum. The first question to ask is, what happens if we try to divide a line segment into smaller and smaller pieces by repeated halving?

Option 1: a line segment cannot be halved more than a certain finite number of times. In this case the continuum has a discrete structure.

Option 2: a line segment can be halved any finite number of times. In this case the next question is, can we think of this endless subdivision process as completable? To state it more precisely, can we survey the whole collection of parts produced by this subdivision?

Option 2.1: the process of repeated subdivision is completable. Then we can take the intersection of a nested infinite sequence of parts, $P_1 \supset P_2 \supset P_3 \supset \cdots$, produced by subdivision, thus forming a smaller part, $\bigcap_{n=1}^{\infty} P_n$, which I shall call an *ω-part*. Since the continuum is homogeneous, the ω-parts are all alike and the original line segment is the union of all its ω-parts. The next question is, is an ω-part itself divisible?

Option 2.1.1: the ω-parts are indivisible and hence are the smallest parts of the continuum. The next question is, what is their length?

Option 2.1.1.1: the ω-parts are of zero length. This is the view taken by conventional modern analysis: the continuum is considered as a set of points, and an ω-part is a singleton set containing one point. It is conventional to add a completeness axiom, stating that any criterion for dividing a line segment into two connected parts determines a point where the division occurs.

Option 2.1.1.2: the ω-parts are of positive length, though still indivisible.

Option 2.1.2: the ω-parts are divisible, so we can continue the process of subdivision. On this view the continuum is represented by a non-archimedean field; an ω-part consists of parts separated by infinitesimal distances. This view was taken by Peirce; he regarded the continuum as an inexhaustible source of parts, exceeding in cardinality any transfinite set [Zink, 2001].

Option 2.2: the process of repeated subdivision is not completable. However many times one has divided one can always divide one step further, but it is illegitimate to consider this unending process as a finished whole. This was Aristotle's view [Lear, 1979–80]. It is also the view that underlies modern 'pointless' topology, in which a space is considered as a lattice of open neighbourhoods rather than as a set of points [Tarski, 1927; Johnstone, 1983; Clarke, 1981, 1985], and also Dummett's [2000] intuitionistic view of physical continua. From this point of view, one admits that a line segment contains points, but only in the 'potential' sense in which a block of marble may be said to contain a statue; a line segment certainly does not *consist of* all its points, any more than a block of marble consists of all its statues.

Zeno's paradoxes were designed to refute each of these options (see Owen [1957–8]), and twentieth-century authors have devised further geometric, kinematic and dynamic paradoxes of infinity [Benardete, 1964; Grünbaum, 1967; Salmon, 1970]. I shall review these paradoxes with a view to discovering their implications for the concept of infinity in the physical world, and in particular for the idea of grounding mathematical infinity in physical models.

4.2 Geometric paradoxes

First let us consider paradoxes relating purely to the structure of space (and matter extended in space), without involving motion. Zeno's paradox of extension was intended to refute the hypothesis that space is infinitely divisible and has indivisible parts (option 2.1.1 above). Under this hypothesis, a line segment consists of infinitely many indivisible parts, all of the same magnitude (by homogeneity). If the parts are of zero magnitude, then the sum of the parts must be zero (however many there are of them); whereas if the parts are of positive magnitude then the sum of them must be infinite. Yet the original line segment was of finite positive magnitude. This seems to rule out options 2.1.1.1 and 2.1.1.2.

The modern analytic theory of the continuum is devised to avoid paradoxes of this sort. This theory adopts option 2.1.1.1, but it also makes a number of distinctions that the ancient Greeks did not. It distinguishes between a line segment and its length; it distinguishes several notions of magnitude (ordinality, cardinality and measure); and it distinguishes several notions of 'sum' (the sum of finitely many numbers, the sum of an infinite series of numbers, and the union of sets). Notice the mixture of potential and actual views of infinity involved here: an infinite sum $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$ does not mean literally the addition of infinitely many numbers but the limit of a succession of finite sums $\frac{1}{2}, \frac{1}{2} + \frac{1}{4}, \frac{1}{2} + \frac{1}{4} + \frac{1}{8}, \dots$ (i.e., potential infinity); whereas an infinite union of sets $\bigcup_{n=1}^{\infty}\left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right] = (0,1]$ is interpreted literally as a straightforward pooling of the contents of all the sets at once (i.e., actual infinity). Having made all these distinctions, we can say that a line segment is the set of all its points and is the union of the singleton sets of its points; however, we can deny that its length is the sum of the lengths of the singleton sets. The line segment has positive length, while the singleton sets have zero length. The length (measure) function μ is countably additive, which means that $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ provided the sets A_n are disjoint and measurable and the infinite series converges. This additivity property does not, however, extend to uncountable sums; indeed, there is no way of adding uncountably many numbers. Hence we cannot form the sum of the lengths of all the singleton sets in a line segment. Zeno's paradox is therefore blocked.

As a piece of pure mathematics this is perfectly consistent; indeed, some authors regard it as a resolution of the paradox [Grünbaum, 1967, chapter III]. However, there are grounds for dissatisfaction with this as a theory of space. Let us consider, for example, *Thomson's Cheese-Grater* [Thomson, 1970]. Compare the following two processes.

(i) Take a lump of cheese and chop it into halves, then chop *one* of the pieces in half, then chop *one* of the pieces so produced in half, and so on.

(ii) Take a lump of cheese and chop it into halves, then chop *both* pieces in half, then chop *all four* pieces in half, and so on.

Process (i) produces a countable infinity of pieces of cheese of positive size, and does not seem particularly paradoxical. But what does (ii) produce? Certainly not chunks of cheese, in the conventional sense. The only thing one can say, if one accepts the conventional theory of the continuum, is that it produces an uncountable infinity of 'cheese-points', dimensionless, indivisible pieces of cheese.

Since both procedures are, in principle, reversible, we ought to be able to stick the pieces together to recover the original lump of cheese. On the other hand, notice that two lumps of cheese of different size, if both subjected to process (ii), will produce identical sets of cheese-points: that is to say, the metric, order, topological and measure properties of the two lumps are broken up by the grating process, leaving just two sets of cheese-points of equal cardinality. Hence it should be possible to grate one lump of cheese by process (ii) and then reassemble it into a lump of a different size.

This of course is reminiscent of the Banach-Tarski paradox: a ball of radius 1 (thought of as a set of points) can be decomposed into six pieces and reassembled by rigid motions into two balls of radius 1 [Banach and Tarski, 1924].

Similar anomalies arise in the attempt to represent physical boundaries [Smith, 1997]. Is the space occupied by a physical body an open set or a closed set? If two bodies are in contact, are the points on the boundary between them occupied by both bodies or neither? If a disc is divided into two halves, one red and one green, what colour is the boundary? One possible answer would be that colour properties only apply to regions, not points; but if the sky varies in colour continuously from blue in the east to white in the west, then it is very natural to model this by assigning a different colour to each point. A mathematician's solution to these problems would be to represent a colour distribution using a function from points to colours, but to regard two such functions as equivalent if they differ only on a set of measure zero, and then to define a colour distribution as an equivalence class of functions rather than as a single function. A similar technique can be applied to the space occupied by bodies. It could be said that this approach amounts to not taking points very seriously.

These issues of spatial modelling are of practical concern to computer scientists in the fields of artificial intelligence, computational linguistics, qualitative spatial reasoning, geographical information systems, and spatial perception [Vieu, 1993; Smith, 1996; Varzi, 1996]. These fields all face the problem of representing and reasoning about real-world objects, masses of liquid, boundaries, holes and shadows, their positions in space and their evolution through time. In a computational setting, finitist representations are forced on us: it is not helpful to represent everything as an uncountable set of points. Mereology is found to be more suitable than set theory, in the spirit of option 2.2.

These examples provide reason for doubting the suitability of the conventional theory of analysis for representing space. As Dummett [2000] complains, the conventional theory provides excess expressive possibilities: it allows one to formulate distinctions and thought experiments that one would prefer to ban on geometric grounds. One may well want to rule out the cheese-grating process (ii), while permitting (i); but on what basis can one justify treating them differently?

The difficulties we have considered in this subsection will recur more forcibly when we proceed to consider kinematic and dynamic paradoxes.

4.3 Kinematic paradoxes

Next let us consider paradoxes involving change and motion (but without considering the forces and dynamical laws that produce motion).

The options listed in §4.1 for the spatial continuum also apply to the temporal continuum. Zeno's paradox of the arrow was directed against the view that time consists of indivisible parts (options 1 and 2.1.1). (I am loosely following the interpretation of Zeno's paradoxes given by Owen [1957–8].) If there are indivisible parts of time (instants), then, during one instant, an arrow in flight cannot move. That is to say, it cannot change its position, since if it did it would be occupying different positions at different parts of the instant, and we would have conceded that the instant was divisible. On the other hand, we are also assuming that time consists entirely of instants, so if the arrow does not move in any instant then it does not move at all.

This paradox is usually resolved by saying that motion is just a matter of being at different positions at different instants of time (this is known as the 'at-at' theory). If we confine our attention to a single instant then the question of motion does not arise: the arrow just has a single position. The concept of motion only applies to intervals of time. Admittedly, we do often speak in physics of the 'instantaneous' velocity of a particle, but this just means the *limit* of its velocity over an interval as the length of the interval tends to zero. This resolution of the paradox seems mathematically and physically satisfactory, provided one has no other objection to indivisible parts of time.

Zeno's paradox of the stadium seems to be directed against the view that space and time consist of indivisible parts of positive length (options 1 and 2.1.1.2). Zeno considers two lines of chariots moving in opposite directions at a speed of one indivisible part of space per indivisible part of time. By considering the relative motion of the two lines, Zeno finds that we are more or less forced to subdivide our parts of space or time. Alternatively one could say that motion that appears smooth at the macroscopic scale consists of a jerky, non-consecutive traversal of positions at the microscopic scale. Since discrete theories of space and time have never been popular I shall not dwell on this paradox further.

More troubling are Zeno's dichotomy paradox and Achilles paradox, which are directed against option 2.2, the view that space and time are infinitely divisible but that the process of infinite subdivision cannot be surveyed as a whole. (These paradoxes also apply against option 2.1.2, at least in its Peircean form in which the process of subdivision is transfinite and incompletable; but I shall stick to option 2.2 for expository convenience.) To traverse a spatial interval $[0,1]$ one must begin by crossing the first half, $[0,\frac{1}{2}]$; but to do so one must begin by crossing the first half of that, $[0,\frac{1}{4}]$, and to do this one must begin by crossing the first half, $[0,\frac{1}{8}]$, and so on. Thus one can never get started. This is the *regressive* version of the dichotomy paradox. The *progressive* version says that to complete the crossing of $[0,1]$ one must cross $[\frac{1}{2},1]$; to do that one must cross $[\frac{3}{4},1]$, and so on. The Achilles paradox involves a race between Achilles and the tortoise, in which Achilles runs ten times as fast as the tortoise but the tortoise has a head start of 1 unit of length. Achilles must traverse the interval $[0,1]$ to reach the tortoise's initial position, by which time the tortoise has reached 1.1; so Achilles must next traverse $[1,1.1]$, by which time the tortoise has reached 1.11; so Achilles must traverse $[1.1,1.11]$; and so on. Thus Achilles must traverse infinitely many intervals to catch up with the tortoise.

In all three cases, a smooth motion is analysed into infinitely many parts by repeated subdivision. Under option 2.2, this subdivision process cannot be thought of as a finished whole; hence the motion can never be completed.

Aristotle believed that this argument could be answered within the terms of option 2.2. The *motion* can be completed but the *subdivision process* cannot. The points of subdivision exist only in a potential sense; hence Achilles does not actually have to traverse infinitely many points [Huggett, 1999, 31-6]. Unfortunately, this solution cannot cope with the 'staccato' version of the paradoxes [Benardete, 1964, 8-9; Grünbaum, 1970]. In the staccato form of the progressive dichotomy, we traverse $[0,\frac{1}{2}]$ in $\frac{1}{4}$ second, then pause for $\frac{1}{4}$ second, then we traverse $[\frac{1}{2},\frac{3}{4}]$ in $\frac{1}{8}$ second, then pause for $\frac{1}{8}$ second, and so on, thus reaching the point 1 after 1 second. In this case the spatial points $\frac{1}{2}, \frac{3}{4}, \ldots$ are individuated by the traversal process, so it cannot be maintained that they are merely potential points.

In fact, this is just one of a long list of modern paradoxes involving 'supertasks', i.e., infinite sets of discrete actions accomplished in a finite time; the idea is that one specifies the individual actions and then enquires what the outcome of the whole process would be. I am only concerned with the conceptual coherence of supertasks here; see Davies [2001] and Cotogno [2003] for discussion of how they might be realised in Newtonian and general relativistic dynamics. Let us first consider supertasks of order-type ω. These involve an infinite sequence of actions $a_1, a_2, a_3, \ldots$, where action a_n is carried out in the time interval $(1-2^{-n+1}, 1-2^{-n})$. We are assuming here that any given action can be carried out at an arbitrarily high speed. All the actions are completed in the time interval $(0, 1)$. The question is, what is the resulting state of affairs at time 1?

The simplest and most telling example is Thomson's lamp [1954–5]. The lamp has a button that switches it from off to on or from on to off. Initially the lamp is off. Action a_n consists of pressing the button. At time 1, is the lamp on or off? There would seem to be five possible answers.

(i) The lamp is on.

(ii) The lamp is off.

(iii) The situation is indeterministic: both outcomes at time 1 are causally possible, given the course of events before time 1.

(iv) The situation is indeterminate: our physical laws break down here and are unable to tell us which outcomes at time 1 are permitted, given the course of events before time 1.

(v) The situation is underdetermined: the outcome depends on other information about the events before time 1 not provided in the description.

Answers (i) and (ii) can be rejected as arbitrary: the problem description is symmetric between on and off, and cannot imply an asymmetric outcome. Answers (iii) and (iv) are very odd, but we shall encounter examples of such unexpected indeterminism and indeterminacy shortly when we come to dynamic paradoxes. Benacerraf [1962], Grünbaum [1967, 92-102] and Harrison [1996] choose answer (v): for them, the outcome depends on the internal mechanism of the lamp switch. Suppose the button controls the position of a lever; if the height of the lever is zero an electrical circuit is made and the lamp is on, whereas if the height is non-zero the circuit is broken and the lamp is off. Consider two possibilities:

a. a 'discontinuous' version, in which the lever rises to a fixed height h whenever the lamp is switched off;

b. a 'diminishing' version, in which, due to the increasing rapidity of switching on and off, the lever rises to decreasing heights h_n (at the nth switching off), where $h_n \to 0$ as $n \to \infty$.

Let $f(t)$ be the position of the lever at time t. The function f is said to be *continuous at 1* iff $f(1) = \lim_{t\to 1} f(t)$. In version (a), f is discontinuous at 1; in fact, the one-sided limit $\lim_{t\to 1^-} f(t)$ does not exist. Some authors [Black, 1950–1; Grünbaum, 1970] dismiss discontinuous trajectories as 'kinematically forbidden', though without giving any justification. A possible *dynamical* motive for insisting on continuity might be a belief that the evolution of a system is governed by differential equations, which would certainly impose smoothness conditions on any dynamically realisable trajectory. If one does not assume the use of differential equations, then on what other basis could the present state of a system determine its future? Is $f(1)$ causally determined by the values of f on $(-\infty, 1)$? If the answer is no then we are back to answers (iii) and (iv).

In version (b), we have a well-defined limit $\lim_{t\to 1^-} f(t) = 0$, so if f is continuous then $f(1) = 0$ and the lamp ends up on.

Many other supertasks of order-type ω also admit a 'discontinuous' version and a 'diminishing' version. The discontinuous version is either banned (if one insists on continuity), indeterministic or indeterminate. The diminishing version has a unique outcome consistent with continuity. For example, Grünbaum [1970] considers both discontinuous and diminishing versions of the staccato dichotomy paradox. Black's (1950–1) infinity machines (which transfer infinitely many marbles from one tray to another) are discontinuous cases, though Grünbaum (1967, p. 103) considers a diminishing version. Benacerraf's (1962) vanishing and shrinking genies also fit into this classification.

We have seen that we can only derive an outcome for a supertask of order-type ω if we are willing to assume continuity (or, to be precise, left-continuity, $f(1) = \lim_{t\to 1^-} f(t)$, where $f(t)$ is the state of the system at time t). Even assuming continuity, however, the outcome can be ambiguous, as for example in various paradoxes involving putting balls into urns [Allis and Koetsier, 1991; Friedman, 2002]. In the Littlewood-Ross paradox, a countable infinity of balls is numbered $1, 2, 3 \ldots$; action a_n consists of putting balls number $10n - 9, \ldots 10n$ into an urn and removing ball number n. At time 1, after all the actions have been carried out, what balls are in the urn? Infinitely many or none of them? Allis and Koetsier [1995] argue, using a continuity principle, that the urn should be empty: their argument amounts to considering the fate of each ball separately on the grounds that each ball is only affected by finitely many actions. On the other hand, if one rubs off the numeric labels on the balls and simply considers the number of balls, $N(t)$, in the urn at time t, one sees that $\lim_{t\to 1^-} N(t) = \infty$, so by continuity $N(1) = \infty$. It beggars belief that the outcome can depend so drastically on the presence of labels on the balls. This discrepancy arises of course because continuity depends on a topology on the set of states; Allis and Koetsier's argument relies on one topology and the opposing argument relies on another. Different topologies are suggested by different ways of describing the supertask and by different assumptions about which aspects are relevant. Hence the outcome is indeterminate until we can justify the choice of one topology.

Let us consider the consequences of this analysis for Russell's and Weyl's notion of an infinite computation, e.g., evaluating Goldbach's conjecture by testing each even number $4, 6, 8, 10, \ldots$ to see whether it is the sum of two primes. When we describe this supertask as a 'computation', we are presupposing two things.

(i) *Determinism*: the outcome of the computation must be determined by the instructions. We must get the same answer every time we repeat it.

(ii) *Abstractness*: the outcome should not depend on whether the computer is made of electronic, mechanical, optical or hydraulic components, but only on the computation performed. An evaluation of Goldbach's conjecture should only depend on arithmetic facts about even numbers and prime numbers, not on the type of hardware, nor the details of the design of the switches and logic gates, nor the colour of the equipment, nor the weather.

If these conditions are not satisfied then we cannot speak of an infinite *computation* but merely an infinite *experiment* and it is no use in elucidating the meaning of an infinitistic statement such as Goldbach's conjecture. Unfortunately, in our analysis we have discovered that there are four possible verdicts on supertasks of order-type ω: (a) banned on kinematic grounds; (b) indeterministic; (c) underdetermined (dependent on the detailed design of switches); and (d) indeterminate (dependent on an arbitrary choice of topology). Hence we do not yet have a convincing example of an infinite computation.

Let us move on to consider supertasks of order-type ω^*. These consist of a reversed infinite sequence of actions $\ldots a_5, a_4, a_3, a_2, a_1$, where a_n is carried out in the time interval $(2^{-n}, 2^{-n+1})$, or possibly (if one allows an infinite past) in the time interval $(-n, -n+1)$. Consider the following puzzle, first published in *Mathematics Magazine* in 1971.

> A boy, a girl, and a dog take a walk along a path; they depart from a common starting point. The boy walks at four miles per hour and the girl walks at three miles per hour; the dog trots back and forth between them at ten miles per hour. Question: at the end of one hour, what is the position of the dog and in which direction is it facing? Answer: the dog can be at any point between the boy and the girl, and it can be facing in either direction. Pick an arbitrary point and either direction. If you time-reverse the whole process, you will find that it invariably leads back to the stated starting conditions. [Salmon, 1970, preface to 2001 reprint, p. ix]

The situation is underdetermined: it needs a boundary condition specifying the dog's position and direction at one point in time. A similar observation applies to other paradoxes of order-type ω^*; for example, in a time-reversed version of Thomson's lamp, the lamp could be on or off after the final action a_1. This is not paradoxical. In the forwards version of Thomson's lamp we expect to have to specify the initial state of the lamp, so in the reversed version we should expect to have to specify the final state.

The example most relevant to our purposes is Benardete's [1964, 123-4] backwards verification of Goldbach's conjecture. The action a_n consists of checking whether $2n + 2$ is the sum of two prime numbers; if it is not then we keep a record of $2n + 2$. At the end we have a list of all counter-examples to Goldbach's conjecture. Unfortunately this procedure is also vulnerable to underdetermination: the list may contain, in addition to the genuine counter-examples we have recorded, spurious numbers that are present because they were there at every stage. Moreover, we cannot check at the end whether these numbers are really counter-examples because there may be infinitely many of them.

Worse still, it is easy to design variations of this backwards computation that have no possible outcome at all: e.g., let a_n consist of testing whether n is a prime number and halting if it is, thus yielding the greatest prime number. This is closely analogous to a series of paradoxes presented by Benardete [1964, 236–239, 252–261]; the general idea of all of them is that an object encounters an ω^*-sequence of barriers, each of which would be sufficient to stop it, but since there is no *first* barrier the object encounters we cannot say which barrier has stopped it. The most troubling version of the paradox involves *non-existent* barriers [259–260]. A man walks along the x-axis towards 0 from the negative direction. A sequence of gods $g_1, g_2, g_3, \ldots$ awaits to put barriers in his path. God g_n resolves to place a barrier at the position 2^{-n} if the man reaches as far as $\frac{3}{4}2^{-n}$. What happens as the man tries to pass through 0? He cannot reach any position beyond 0, since that would entail passing some barrier; so he never gets further than 0. But, in that case, none of the barriers is ever placed. So he is stopped by the gods' *unrealised intentions* to place barriers!

The man's progress is not blocked by a 'strange field of force' or 'before-effect' as suggested by Benardete and some later commentators. Indeed, as Yablo points out, the paradox does not even depend on the presence of the man [2000]. Let B_n be the proposition that god g_n places his barrier (at time 2^{-n}, say). Let I_n be the proposition $B_n \Leftrightarrow \forall m > n \, \neg B_m$, which expresses g_n's intention to place a barrier iff no earlier god has done so. In formal logical terms, the set of propositions $\{I_1, I_2, I_3, \ldots\}$ is consistent but not ω-consistent and not satisfiable (i.e., not satisfiable by any standard model, in which the $\forall m$ quantifier ranges over the standard natural numbers). Hence Yablo is right to say that the gods are prevented from carrying out their intentions by pure logic. No contradiction can be deduced from the gods' intentions; no obstacle prevents any god from carrying out his intention; yet no state of affairs counts as all the gods fulfilling their intentions.

We can dispel the sense of mystery by viewing the matter in the following light: whenever we set up a supertask by specifying what is to happen at each step we are defining the evolution of the system by *recursion* on the steps. As is well known, recursive definitions work perfectly on ω and other well-founded orderings, but fail on non-well-founded orderings, such as ω^*, for which there may be zero, one or more solutions of the recursion equations – exactly as we found in our examples.

This makes our paradoxes look a lot less paradoxical: this failure of recursion is in fact a perfectly familiar and well-understood mathematical phenomenon. This provides a strong reason for dismissing all supertasks of non-well-founded order-type as ill-defined.

Indeed, this rather undermines the idealised world-picture we have been working with. We have assumed that we are free to imagine an infinity of objects, to assign to them arbitrary properties, to arrange them in any non-overlapping configuration, to assume that any 'god' or machine (or whatever) will carry out its instructions correctly, and then to ask what happens when the set-up is left to run. We have found that in some cases the set-up simply *cannot* evolve and hence cannot exist in a world that has time. Nevertheless I shall not abandon the idealised world-picture; I shall persevere with it, assuming that it is well-behaved as long as we do not try to set up temporally non-well-founded systems of events.

4.4 Paradoxes involving sums of conserved quantities

It is a standard result of elementary analysis that the infinite series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ can be made to converge to any sum, or to diverge, by reordering the terms suitably [Haggarty, 1993, §4.2]. This can be turned into a physical paradox if we imagine a countably infinite set of particles $p_1, p_2, \ldots$, where p_n has an electric charge (or any other conserved quantity) of $\frac{(-1)^n}{n}$. Suppose the particles collide simultaneously and coalesce. What is the charge of the resulting particle? The problem is that infinite sums are defined only for *sequences* of numbers, whereas the physical situation contains an *unordered set* of particles.

This type of paradox was introduced by Cooke, using examples with particles of unbounded density, unbounded speed or unbounded length (2003). If we are willing to assume an upper bound ρ on (positive or negative) charge density, an upper bound v on speed, and that the resulting composite particle is bounded by a sphere of radius r, then we can exclude the paradox as follows. Let us generalise the situation to allow the particles to have any charges $q_1, q_2, \ldots$, and let us suppose that the particles coalesce within a time period T. Then at the beginning of the time period they must all have been bounded by a sphere of radius $r + vT$, and hence $\sum_{n=1}^{\infty} |q_n| \leq \frac{4}{3}\pi(r+vT)^3\rho$. Thus the sequence (q_n) is absolutely convergent, and hence (by a standard theorem) its sum $\sum_{n=1}^{\infty} q_n$ converges and is independent of the order of terms. In particular, coalescence in a finite time is impossible in the case $q_n = \frac{(-1)^n}{n}$.

Unfortunately, Cooke's paper also shows how to realise the summation $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ as a sum of gravitational forces, using a static configuration of particles of equal mass and volume; this example cannot be excluded by limits on density, speed or size. Cooke suggests that there may be some 'intrinsic ordering' hidden in the physical situation that specifies how the terms are to be summed, but it is very hard to see where such an ordering could be found.

There is a second type of problematic infinite summation that occurs in physics, where a sum or limit of positive quantities diverges unexpectedly in what seems to be a feasible physical situation. In classical physics the electromagnetic or gravitational self-energy of a point particle diverges (that is, it tends to infinity like $\frac{1}{r}$ as the radius r tends to 0, for fixed charge or mass). There is also Olbers' paradox of conservation of flux [Wesson, 1991]: if an infinite static universe is sprinkled evenly with stars of uniform luminosity then the light intensity at any point tends to infinity over time.

These paradoxes provide some motivation for abandoning the 'limit of partial sums' view of infinite summation and re-asserting a naive 'actual infinity' view, as suggested in §1, under which infinitely many quantities (of the same qualitative kind) can be simply lumped together into a total quantity, in precisely the same way that finitely many quantities can (cf. [Benardete, 1964, 26]). This entails accepting the possibility of infinite quantities, of course. The $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ problem can be dismissed by rejecting the concept of negative quantities. This view fits quite well with the way we naively think of quantities: intuitively, a quantity (e.g., the lifetime of the universe) can be finite or infinite but cannot be negative. Negative quantities are merely a figure of speech: a 'negative acceleration' is not really an acceleration at all but a deceleration; a 'negative profit' is not really a profit but a loss. Profits and losses differ qualitatively (not merely in sign), so, although it is always possible to add infinitely many profits together to give a (finite or infinite) total, we cannot be expected to add infinitely many profits and losses together. This view merits further development; notice, however, that it involves unpicking the laboriously constructed solution to Zeno's paradox of extension adopted by conventional mathematics (see §4.2).

4.5 Paradoxes of collision

Consider an infinite collection of bodies that interact only by elastic collision, according to Newton's three laws of motion. Assume that the bodies can be made as nearly rigid as required: i.e., any collision can be assumed to be completed in an arbitrarily small time. Newtonian mechanics is commonly supposed to be deterministic and to satisfy laws of conservation of mass, momentum and energy, but there are four ways in which these properties can be subverted.

The first method is by escape to infinity. Laraudogoitia [1997] describes how, by an ω-sequence of elastic collisions, a particle can be made to double its speed at ever-halving intervals and so escape to spatial infinity in a finite time (like Benardete's metaphysical rocket [1964, 149]). Hence the particle disappears from space (and in fact so do all the other particles involved in the collisions). This phenomenon of escape to infinity is a genuine feature of Newtonian mechanics, not an artifact of our unrealistic assumptions of infinitely many objects and perfectly elastic collisions: Xia [1992] has shown that five point-mass particles interacting via Newton's law of gravity can escape to spatial infinity in a finite time without colliding. Moreover, since Newton's laws are time-reversible, it is equally possible for particles to appear spontaneously from infinity, in an ω^*-sequence of collisions. We know from §4.3 that ω^*-sequences of actions can be indeterministic, so perhaps this is no great surprise. (There are also cases in which collision set-ups of order-type ω^* have no solutions at all, such as the one described by Angel [2001], but I shall not consider this phenomenon as it has already been dealt with in §4.3.)

The second method is by using infinite mass to evade the conservation laws. Consider infinitely many balls $B_1, B_2, B_3, \ldots$ of equal mass, where B_n has radius 2^{-n}. We can enclose them all in a spherical shell to produce a composite body of infinite mass and finite volume. Take two such bodies and collide them; what happens? The result is indeterminate, since total mass and momentum are undefined, so the law of conservation of momentum cannot be applied.

The third method involves momentum disappearing into an ω-sequence. Laraudogoitia [1996; 1999] imagines a moving particle P_0 colliding elastically with a sequence of stationary particles $P_1, P_2, P_3, \ldots$, all of equal mass. In accordance with Newton's laws, P_0 passes all its momentum on to P_1, which then passes all its momentum on to P_2, and so on. After a finite time all the collisions are complete and all the particles are stationary. The momentum has disappeared. In the time-reversed version motion appears spontaneously from stationary particles.

Laraudogoitia's [2002] example of spontaneous self-excitation exploits both the second and third methods.

The fourth method is by imitating Thomson's lamp (see §4.3): we set up an oscillating ω-sequence of collisions that has no possible outcome compatible with continuity [Earman and Norton, 1998; Laraudogoitia, 1998]. Newton's laws are indeterminate in such a case. Laraudogoitia suggests that the particles simply disappear when their world-lines cannot be continued in a continuous way. He also believes that the situation is time-reversible, so that particles can appear unpredictably out of the void.

Perhaps it is time to take a closer look at conservation laws. Earman [1986] points out that they can be formulated in a variety of ways. In most of our examples momentum is conserved *locally* (at each collision) but not *globally* (the total momentum of the system is not equal at all times). The relation between the local and global conservation laws can be most easily seen from field theory [Weinberg, 1995, §7.3], since any situation involving particles can be modelled in terms of density fields. A continuous symmetry of the action implies a local conservation law

$$\frac{\partial J^\mu(x)}{\partial x^\mu} = 0$$

where J is a four-vector current field, $x = (t, \mathbf{x})$ is a point in space-time, and μ is a Lorentz index. Using Stokes' theorem we can deduce the global conservation law

$$\frac{dQ(t)}{dt} = 0, \quad \text{where } Q(t) = \int J^0(t, \mathbf{x})\, d^3\mathbf{x}.$$

This argument assumes that J vanishes sufficiently fast at large spatial distances. It also presupposes that J is differentiable. These two conditions rule out *all* the above cases. Any case involving escape to infinity or particles spread out over unbounded space violates the condition of vanishing at infinity. Any case involving infinitely many equal masses in a bounded space-time volume violates the differentiability condition, since, by the properties of compact sets, there must exist a point every neighbourhood of which contains infinite mass.

Hence, in mathematical terms, there is nothing paradoxical about the above cases: they simply demonstrate that conservation laws are subject to conditions of smoothness and vanishing at infinity. However, they are certainly paradoxical physically. After all, the purpose of conservation laws is to explain the stability of the world; if the laws are formulated in a weakened form that permits matter and motion to appear and disappear spontaneously (as advocated by Earman [1986] and Laraudogoitia [1998]) then they fail to serve their purpose and we need *another* physical principle to explain why this never seems to happen in nature.

4.6 Conclusions: does nature abhor infinities?

It is very natural to assume that the world is an infinitely divisible space-time manifold containing a possibly infinite number of non-overlapping particles with arbitrary values for speed, mass, etc.; we use this picture as the starting point for our physical theorising and as an aid to thinking about actual infinity in pure mathematics. All the physical paradoxes we have considered involve problems with this idealised picture or a mismatch between it and the mathematical systems we use in our physical theories. None of the paradoxes reveals any inconsistency or conceptual incoherence in the mathematical systems considered as pure mathematics. Equally, none of them would be troubling to a physicist. Worse things happen in physics every day. Physicists are used to calculating with infinite and infinitesimal quantities to obtain finite results, most notoriously in the divergent integrals and divergent perturbation series that occur in quantum field theory; they think nothing of treating space-time as $(4+\epsilon)$-dimensional, where ϵ is a small complex number, rather than 4-dimensional, if that is what it takes to make their integrals converge (this technique is known as dimensional regularisation [Weinberg, 1995, §11.2]). Physicists are tolerant of infinities appearing in their calculations because they regard their theories merely as low-energy approximations to some more comprehensive theory in which the calculation would converge [Weinberg, 1995, §12.3]. The infinities in quantum field theory are believed to be ultimately due to the use of local field operators, which embody the idealised view that particles and their interactions can be localised to a point [Cao and Schweber, 1993]. More generally, infinities arise in calculations due to unrealistic simplifying assumptions and they are dealt with by modifying the assumptions.

Nevertheless, the paradoxes are problematic from the point of view of this essay, namely the attempt to use physical models to demonstrate the conceptual coherence of actual infinity, in order to justify its use in mathematics. The geometric paradoxes cast doubt on the view that space is an infinite set of points; the kinematic paradoxes undermine the notion of an infinite computation; and the dynamic paradoxes undermine the notion of an infinity of objects. The ω^*-sequence paradoxes discussed at the end of §4.3 are particularly troubling. If the world could contain infinitely many objects, what is to stop them from arranging themselves into a configuration in which they cannot evolve?

One possible reaction to the paradoxes is to impose boundary conditions on our idealised picture, such as banning infinite numbers of objects, imposing a bound on speed and density, or insisting that matter is represented by fields that are differentiable and vanish at infinity. This would certainly tame the paradoxes but is hard to motivate on *a priori* grounds.

A second possible reaction is to reject the idealised picture as incoherent and insist that we stick to real physics, in which paradoxes are excluded by the speed of light limitation, the atomic structure of matter, the uncertainty principle, or some other limiting principle. Note that each such limiting principle is merely contingent, but it may be a necessary truth that *some* limiting principle exist to cover each paradox.

Certainly, the idea of a world of infinitely many objects is not an innocent and straightforward extension of the idea of a world of finitely many objects. Unexpected global difficulties arise when one passes from finite to infinite, and they are different from the difficulties that arise for infinity in pure mathematics. That is not to say that these difficulties are insuperable, merely that supporters of actual infinity have a lot of explaining to do.

BIBLIOGRAPHY

[Aczel, 1988] P. Aczel. *Non-well-founded Sets.* Stanford: CSLI Publications, 1988.

[Allis and Koetsier, 1991] V. Allis and T. Koetsier. On some paradoxes of the infinite. *British Journal for the Philosophy of Science*, **42**, no. 2, 187–94, 1991.

[Allis and Loetsier, 1995] V. Allis and T. Koetsier. On some paradoxes of the infinite II. *British Journal for the Philosophy of Science*, **46**, no. 2, 235–47, 1995.

[Angel, 2001] L. Angel. A physical model of Zeno's dichotomy. *British Journal for the Philosophy of Science*, **52**, no. 2, 347–58, 2001.

[Baire, 1905] R. Baire. Letter to J. Hadamard. Translated by G.H. Moore. In Ewald (1996), vol. 2, pp. 1079–80, 1905.

[Balaguer, 1998] M. Balaguer. *Platonism and Anti-Platonism in Mathematics.* Oxford: Oxford University Press, 1998.

[Banach and Tarski, 1924] S. Banach and A. Tarski. Sur la décomposition des ensembles de points en parties respectivement congruentes. *Fundamenta Mathematicae*, **6**, 244–77, 1924.

[Beeson, 1985] M. J. Beeson. *Foundations of Constructive Mathematics: Metamathematical Studies.* Berlin: Springer-Verlag, 1985.

[Bell, 1981] J. L. Bell. Category theory and the foundations of mathematics. *British Journal for the Philosophy of Science*, **32**, no. 4, 349–58, 1981.

[Bell, 1998] J. L. Bell. *A Primer of Infinitesimal Analysis.* Cambridge: Cambridge University Press, 1998.

[Benacerraf, 1962] P. Benacerraf. Tasks, super-tasks, and the modern Eleatics. *Journal of Philosophy*, **LIX**, 765–84. Reprinted in Salmon (1970), pp. 103–129, 1962.

[Benacerraf, 1973] P. Benacerraf. Mathematical truth. *The Journal of Philosophy*, **70**, no. 19, 661–79, 1973.

[Benacerraf and Putnam, 1983] P. Benacerraf and H. Putnam, eds. *Philosophy of Mathematics.* 2nd edition. Cambridge: Cambridge University Press, 1983.

[Benardete, 1964] J. A. Benardete. *Infinity: An Essay in Metaphysics.* Oxford: Clarendon Press, 1964.

[Bernays, 1935] P. Bernays. On platonism in mathematics. Translated by C.D. Parsons. In [Benacerraf and Putnam, 1983, 258–71].

[Bishop, 1967] E. Bishop. *Foundations of Constructive Analysis.* New York: McGraw-Hill, 1967.

[Bishop, 1970] E. Bishop. Mathematics as a numerical language. In A. Kino, J. Myhill and R.E. Vesley (eds) *Intuitionism and Proof Theory.* Proceedings of the summer conference at Buffalo, New York, 1968. Amsterdam: North-Holland, pp. 53–71, 1970.

[Black, 1950–1] M. Black. Achilles and the tortoise. *Analysis*, **XI**, 91–101, 1950–1. Reprinted in [Salmon, 1970, 67-81].

[Bolzano, 1810] B. Bolzano. Contributions to a better-grounded presentation of mathematics, 1810. Translated by S. Russ. In [Ewald, 1996, vol. 1, 174–224].

[Bolzano, 1817] B. Bolzano. Purely analytic proof of the theorem that between any two values which give results of opposite sign there lies at least one real root of the equation, 1817. Translated by S. Russ. In [Ewald, 1996, vol. 1, 225–48].

[Bolzano, 1851] B. Bolzano. *Paradoxes of the Infinite*, 1851. Translated by D.A. Steele, 1950. London: Routledge & Kegan Paul.

[Boolos, 1971] G. Boolos. The iterative conception of set. *Journal of Philosophy*, **68**, 215–32, 1971. Reprinted in [Benacerraf and Putnam, 1983, 486-502].

[Bridges and Richman, 1987] D. Bridges and F. Richman. *Varieties of Constructive Mathematics.* Cambridge: Cambridge University Press, 1987.

[Brouwer, 1907] L. E. J. Brouwer. On the Foundations of Mathematics, 1907. In A. Heyting (ed.) (1975) *L.E.J. Brouwer: Collected Works.* Vol. I. Amsterdam: North-Holland, pp. 11–101.

[Cantor, 1874] G. Cantor. On a property of the set of real algebraic numbers, 1874. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 839–43].

[Cantor, 1883] G. Cantor. Foundations of a general theory of manifolds: a mathematico-philosophical investigation into the theory of the infinite, 1883. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 878–920].

[Cantor, 1885] G. Cantor. Rezension der Schrift von G. Frege 'Die Grundlagen der Arithmetik'. *Deutsche Literaturzeitung*, **6**, 728–9, 1885.

[Cantor, 1891] G. Cantor. On an elementary question in the theory of manifolds, 1891. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 920–2].

[Cantor, 1899a] G. Cantor. Letter to R. Dedekind, 3rd August 1899. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 931–5].

[Cantor, 1899b] G. Cantor. Letter to R. Dedekind, 28th August 1899. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 936–7].

[Cantor, 1899c] G. Cantor. Letter to R. Dedekind, 31st August 1899. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 939–40].

[Cao and Schweber, 1993] T. Y. Cao and S. S. Schweber. The conceptual foundations and the philosophical aspects of renormalization theory. *Synthese*, **97**, no. 1, 33–108, 1993.

[Chomsky, 1972] N. Chomsky. *Language and Mind.* Enlarged Edition. New York: Harcourt Brace Jovanovich, 1972.

[Clarke, 1981] B. L. Clarke. A calculus of individuals based on connection. *Notre Dame Journal of Formal Logic*, **22**, no. 3, 204–18, 1981.

[Clarke, 1985] B. L. Clarke. Individuals and points. *Notre Dame Journal of Formal Logic*, **26**, no. 1, 61–75, 1985.

[Cohen, 1966] P. J. Cohen. *Set Theory and the Continuum Hypothesis.* New York: W.A. Benjamin, 1966.

[Cooke, 2003] M. C. Cooke. Infinite sequences: finitist consequence. *British Journal for the Philosophy of Science*, **54**, 591–9, 2003.

[Cotogno, 2003] P. Cotogno. Hypercomputation and the physical Church-Turing thesis. *British Journal for the Philosophy of Science*, **54**, 181–223, 2003.

[Dauben, 1979] J. W. Dauben. *Georg Cantor: His Mathematics and Philosophy of the Infinite.* Cambridge, Massachusetts: Harvard University Press, 1979.

[Davies, 2001] E. B. Davies. Building infinite machines. *British Journal for the Philosophy of Science*, **52**, 671–82, 2001.

[Dedekind, 1872] R. Dedekind. Continuity and irrational numbers, 1872. Translated by W.W. Beman, with revisions by W. Ewald. In [Ewald, 1996, vol. 2, 765–79].

[Dedekind, 1888a] R. Dedekind. Letter to H. Weber, 1888. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 834–5].

[Dedekind, 1888b] R. Dedekind. Was sind und was sollen die Zahlen? 1888. Translated by W.W. Beman, with revisions by W. Ewald. In [Ewald, 1996, vol. 2, 787–833].

[Dedekind, 1890] R. Dedekind. Letter to H. Keferstein, 1890. Translated by H. Wang and S. Bauer-Mengelberg. In [van Heijenoort, 1967, 98–103].

[Díez, 2000] G. F. Díez. Five observations concerning the intended meaning of the intuitionistic logical constants. *Journal of Philosophical Logic*, **29**, no. 4, 409–24, 2000.

[Dummett, 1975a] M. Dummett. Wang's paradox. *Synthese*, **30**, 301–24, 1975.

[Dummett, 1975b] M. Dummett. The philosophical basis of intuitionistic logic. In H.E. Rose and J.C. Shepherdson (eds) *Logic Colloquium '73 (Bristol).* Amsterdam: North-Holland, pp. 5–40, 1975.

[Dummett, 1977] M. Dummett. *Elements of Intuitionism.* Oxford: Clarendon Press, 1977.

[Dummett, 1994] M. Dummett. What is mathematics about? In A. George (ed.) *Mathematics and Mind.* Oxford: Oxford University Press, pp. 11–26, 1994.

[Dummett, 2000] M. Dummett. Is time a continuum of instants? *Philosophy*, **75**, no. 294, 497–515, 2000.

[Earman, 1986] J. Earman. *A Primer on Determinism.* Dordrecht: Reidel, 1986.

[Earman and Norton, 1998] J. Earman and J. D. Norton. Comments on Laraudogoitia's 'Classical particle dynamics, indeterminism and a supertask'. *British Journal for the Philosophy of Science*, **49**, no. 1, 123–33, 1998.

[Ewald, 1996] W. Ewald, ed. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics.* 2 vols. Oxford: Clarendon Press, 1996.

[Feferman, 1998] S. Feferman. *In the Light of Logic.* New York: Oxford University Press, 1998.

[Field, 1980] H. H. Field. *Science without Numbers: A Defence of Nominalism.* Oxford: Blackwell, 1980.

[Fletcher, 1998] P. Fletcher. *Truth, Proof and Infinity: A Theory of Constructions and Constructive Reasoning.* Synthese Library, vol. 276. Dordrecht: Kluwer, 1998.

[Fletcher, 2002] P. Fletcher. A constructivist perspective on physics. *Philosophia Mathematica* (3), **10**, no. 1, 26–42, 2002.

[Fodor and Pylyshyn, 1988] J. A. Fodor and Z. W. Pylyshyn. Connectionism and cognitive architecture: a critical analysis. *Cognition*, **28**, nos 1 & 2, 3–71, 1988.

[Frege, 1879] G. Frege. Begriffsschrift, a formula language, modelled upon that of arithmetic, for pure thought, 1879. Translated by S. Bauer-Mengelberg. In [van Heijenoort, 1967, 1–82].

[Frege, 1884] G. Frege. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, 1884. Translated by J.L. Austin, 2nd revised edition, 1953. Oxford: Blackwell.

[Frege, 1891] G. Frege. Function and concept. Translated by P. Geach. In B. McGuinness (ed.) *Gottlob Frege: Collected Papers on Mathematics, Logic, and Philosophy.* Oxford: Blackwell, pp. 137–156, 1891.

[Frege, 1893] G. Frege. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet.* Vol. 1. Jena: Verlag Hermann Pohle, 1893. English translation of §§0–52 by M. Furth (1967) *The Basic Laws of Arithmetic: Exposition of the System.* Berkeley: University of California Press.

[Frege, 1919] G. Frege. Notes for Ludwig Darmstaedter, 1919. In H. Hermes, F. Kambartel & F. Kaulbach (eds) (1979) *Gottlob Frege: Posthumous Writings.* Oxford: Blackwell, pp. 253–7.

[Friedman, 2002] K. S. Friedman. A small infinite puzzle. *Analysis*, **62**, no. 4, 344–5, 2002.

[George, 1988] A. George. The conveyability of intuitionism, an essay on mathematical cognition. *Journal of Philosophical Logic*, **17**, no. 2, 133–56, 1988.

[Gödel, 1933] K. Gödel. The present situation in the foundations of mathematics, 1933. In S. Feferman (ed.) (1995) *Kurt Gödel: Collected Works.* Vol. III. New York: Oxford University Press, pp. 45–53.

[Gödel, 1938] K. Gödel. Lecture at Zilsel's, 1938. In S. Feferman (ed.) (1995) *Kurt Gödel: Collected Works.* Vol. III. New York: Oxford University Press, pp. 86–113.

[Gödel, 1941] K. Gödel. In what sense is intuitionistic logic constructive?, 1941. In S. Feferman (ed.) (1995) *Kurt Gödel: Collected Works.* Vol. III. New York: Oxford University Press, pp. 189–200.

[Gödel, 1958] K. Gödel. On a hitherto unutilized extension of the finitary standpoint, 1958. Translated by S. Bauer-Mengelberg & J. van Heijenoort. In S. Feferman (ed.) (1990) *Kurt Gödel: Collected Works.* Vol. II. New York: Oxford University Press, pp. 217–51.

[Gödel, 1983] K. Gödel. What is Cantor's continuum problem? In [Benacerraf and Putnam, 1983, 470–85].

[Goodman, 1970] N. D. Goodman. A theory of constructions equivalent to arithmetic. In A. Kino, J. Myhill and R.E. Vesley (eds) *Intuitionism and Proof Theory.* Proceedings of the summer conference at Buffalo, New York, 1968. Amsterdam: North-Holland, pp. 101–20, 1970.

[Goodman, 1973] N. D. Goodman. The arithmetic theory of constructions. In A.R.D. Mathias & H. Rogers (eds) *Cambridge Summer School in Mathematical Logic (Cambridge, 1971).* Lecture Notes in Mathematics, vol. 337. Berlin: Springer-Verlag, pp. 274–98, 1973.

[Grünbaum, 1967] A. Grünbaum. *Modern Science and Zeno's Paradoxes.* London: Allen & Unwin, 1967.

[Grünbaum, 1970] A. Grünbaum. Modern science and Zeno's paradoxes of motion. In Salmon (1970), pp. 200–50, 1970.

[Haggarty, 1993] R. Haggarty. *Fundamentals of Mathematical Analysis.* 2nd edition. Wokingham, U.K.: Addison-Wesley, 1993.

[Hallett, 1984] M. Hallett. *Cantorian Set Theory and Limitation of Size.* Oxford: Clarendon Press, 1984.

[Harrison, 1996] C. Harrison. The three arrows of Zeno: Cantorian and non-Cantorian concepts of the continuum and of motion. *Synthese*, **107**, no. 2, 271–92, 1996.

[Hellman, 1989] G. Hellman. *Mathematics without Numbers: Towards a Modal-Structural Interpretation.* Oxford: Clarendon Press, 1989.

[Hellman, 1998] G. Hellman. Mathematical constructivism in spacetime. *British Journal for the Philosophy of Science*, **49**, no. 3, 425–50, 1998.

[Herbrand, 1931] J. Herbrand. On the consistency of arithmetic, 1931. Translated by J. van Heijenoort. In [van Heijenoort, 1967, 618–28].

[Heyting, 1931] A. Heyting. The intuitionist foundations of mathematics, 1931. Translated by E. Putnam and G.J. Massey. In [Benacerraf and Putnam, 1983, 52–61].

[Heyting, 1956] A. Heyting. *Intuitionism: An Introduction.* Amsterdam: North-Holland, 1956.

[Hilbert, 1900] D. Hilbert. Mathematical problems. Translated by M.W. Newson. In *Bulletin of the American Mathematical Society* (2), **8**, 1902, 437–79, 1900.

[Hilbert, 1925] D. Hilbert. On the infinite, 1925. Translated by S. Bauer-Mengelberg. In [van Heijenoort, 1967, 367–92].

[Hilbert, 1927] D. Hilbert. The foundations of mathematics, 1927. Translated by S. Bauer-Mengelberg and D. Føllesdal. In [van Heijenoort, 1967, 464–79].

[Hilbert, 1930] D. Hilbert. Logic and the knowledge of nature, 1930. Translated by W. Ewald. In [Ewald, 1996, vol. 2, pp. 1157–65].

[Hilbert, 1931] D. Hilbert. The grounding of elementary number theory, 1931. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 1148–57].

[Huggett, 1999] N. Huggett, ed. *Space from Zeno to Einstein: Classic Readings with a Contemporary Commentary.* Cambridge, Massachusetts: MIT Press, 1999.

[Johnstone, 1983] P. T. Johnstone. The point of pointless topology. *Bulletin (New Series) of the American Mathematical Society*, **8**, no. 1, 41–53, 1983.

[Kleene, 1945] S. C. Kleene. On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic*, **10**, no. 4, 109–24, 1945.

[Kock, 1981] A. Kock. *Synthetic Differential Geometry.* Cambridge: Cambridge University Press, 1981.

[Kolmogorov, 1932] A. N. Kolmogorov. On the interpretation of intuitionistic logic, 1932. Translated by P. Mancosu. In [Mancosu, 1998, 328–34].

[Kreisel, 1962] G. Kreisel. Foundations of intuitionistic logic. In E. Nagel, P. Suppes and A. Tarski (eds) *Logic, Methodology and Philosophy of Science.* Stanford: Stanford University Press, pp. 198–210, 1962.

[Kreisel, 1965] G. Kreisel. Mathematical logic. In T.L. Saaty (ed.) *Lectures on Modern Mathematics.* Vol. III. New York: Wiley, pp. 95–195, 1965.

[Lake, 1979] J. Lake. The approaches to set theory. *Notre Dame Journal of Formal Logic*, **20**, no. 2, 415–37, 1979.

[Lear, 1979–80] J. Lear. Aristotelian infinity. *Proceedings of the Aristotelian Society* (new series), **80**, 187–210, 1979–80.

[Lewis, 1991] D. Lewis. *Parts of Classes.* Oxford: Blackwell, 1991.

[Maddy, 1988] P. Maddy. Believing the axioms, I. *Journal of Symbolic Logic*, **53**, no. 2, 481–511, 1988.

[Maddy, 1990] P. Maddy. *Realism in Mathematics.* Oxford: Clarendon Press, 1990.

[Maddy, 1997] P. Maddy. *Naturalism in Mathematics.* Oxford: Clarendon Press, 1997.

[Mancosu, 1998] P. Mancosu, ed. *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s.* Oxford: Oxford University Press, 1998.

[Markov, 1968] A. A. Markov. An approach to constructive mathematical logic. In B. van Rootselaar and J.F. Staal (eds) *Logic, Methodology and Philosophy of Science III.* Proceedings of the Third International Congress for Logic, Methodology and Philosophy of Science, Amsterdam, 1967. Amsterdam: North-Holland, pp. 283–94, 1968.

[Martin-Löf, 1975] P. Martin-Löf. An intuitionistic theory of types: predicative part. In H.E. Rose and J.C. Shepherdson (eds) *Logic Colloquium '73 (Bristol).* Amsterdam: North-Holland, pp. 73–118, 1975.

[Martin-Löf, 1984] P. Martin-Löf. *Intuitionistic Type Theory.* Napoli: Bibliopolis, 1984.

[Mayberry, 1994] J. P. Mayberry. What is required of a foundation for mathematics? *Philosophia Mathematica* (3), **2**, no. 1, 16–35, 1994.

[Mayberry, 2000] J. P. Mayberry. *The Foundations of Mathematics in the Theory of Sets.* Cambridge: Cambridge University Press, 2000.

[Mirimanoff, 1917] D. Mirimanoff. Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles. *L'Enseignement Mathématique*, **19**, 37–52, 1917.

[Moore, 1982] G. H. Moore. *Zermelo's Axiom of Choice: Its Origins, Development, and Influence.* New York: Springer-Verlag, 1982.

[Owen, 1957–8] G. E. L. Owen. Zeno and the mathematicians. *Proceedings of the Aristotelian Society* (new series), **58**, 199–222. Reprinted in Salmon (1970), pp 139-163. 1957–8.

[Parikh, 1971] R. Parikh. Existence and feasibility in arithmetic. *Journal of Symbolic Logic*, **36**, no. 3, 494–508, 1971.

[Parsons, 1977] C. Parsons. What is the iterative conception of set? In R.E. Butts and J. Hintikka (eds) *Proceedings of the 5th International Congress of Logic, Methodology and Philosophy of Science 1975.* Dordrecht: Reidel, pp. 335–67, 1977.

[Peirce, 1885] C. S. Peirce. On the algebra of logic: a contribution to the philosophy of notation. *American Journal of Mathematics*, **7**, 180–202, 1885. Reprinted in [Ewald, 1996, vol. 1, 608-32].

[Peressini, 1999] A. Peressini. Applying pure mathematics. *Philosophy of Science*, **66**, no. 3, S1–S13, 1999.

[Pérez Laraudogoitia, 1996] J. Pérez Laraudogoitia. A beautiful supertask. *Mind*, **105**, no. 417, 81–3, 1996.

[Pérez Laraudogoitia, 1997] J. Pérez Laraudogoitia. Classical particle dynamics, indeterminism and a supertask. *British Journal for the Philosophy of Science*, **48**, no. 1, 49–54, 1997.

[Pérez Laraudogoitia, 1998] J. Pérez Laraudogoitia. Infinity machines and creation ex nihilo. *Synthese*, **115**, no. 2, 259–65, 1998.

[Pérez Laraudogoitia, 1999] J. Pérez Laraudogoitia. Earman and Norton on supertasks that generate indeterminism. *British Journal for the Philosophy of Science*, **50**, no. 1, 137–41, 1999.

[Pérez Laraudogoitia, 2002] J. Pérez Laraudogoitia. Just as beautiful but not (necessarily) a supertask. *Mind*, **111**, no. 442, 281–7, 2002.

[Placek, 1999] T. Placek. *Mathematical Intuitionism and Intersubjectivity: A Critical Exposition of Arguments for Intuitionism.* Synthese Library, vol. 279. Dordrecht: Kluwer, 1999.

[Poincaré, 1905] J. H. Poincaré. Mathematics and logic: I, 1905. Revised in 1908. Translated by G.B. Halsted and W. Ewald. In [Ewald, 1996, vol. 2, 1021–38].

[Poincaré, 1906a] J. H. Poincaré. Mathematics and logic: II, 1906. Translated by G.B. Halsted and W. Ewald. In [Ewald, 1996, vol. 2, 1038–52].

[Poincaré, 1906b] J. H. Poincaré. Mathematics and logic: III, 1906. Translated by G.B. Halsted and W. Ewald. In [Ewald. 1996, vol. 2, 1052–71].

[Poincaré, 1910] J. H. Poincaré. On transfinite numbers, 1910. Translated by W. Ewald. In [Ewald, 1996, vol. 2, 1071–4].

[Poincaré, 1913] J. H. Poincaré. *Mathematics and Science: Last Essays*, 1913. Translated by J.W. Bolduc, 1963. New York: Dover.

[Pollard, 1990] S. Pollard. *Philosophical Introduction to Set Theory.* Notre Dame, Indiana: University of Notre Dame Press, 1990.

[Pollard, 1996] S. Pollard. Sets, wholes, and limited pluralities. *Philosophia Mathematica* (3), **4**, no. 1, 42–58, 1996.

[Pozsgay, 1971] L. Pozsgay. Liberal intuitionism as a basis for set theory. In D.S. Scott (ed.) *Axiomatic Set Theory, I.* Providence, Rhode Island: American Mathematical Society, pp. 321–30, 1971.

[Pudlák and Sochor, 1984] P. Pudlák and A. Sochor. Models of the alternative set theory. *Journal of Symbolic Logic*, **49**, no. 2, 570–85, 1984.

[Quine, 1937] W. V. O. Quine. New foundations for mathematical logic. *American Mathematical Monthly*, **44**, no. 2, 70–80, 1937.

[Robinson, 1966] A. Robinson. *Non-standard Analysis.* Amsterdam: North-Holland, 1966.

[Russell, 1903] B. Russell. *The Principles of Mathematics.* London: Allen and Unwin, 1903.

[Russell, 1906] B. Russell. On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society* (2), **4**, 29–53, 1906.

[Russell, 1908] B. Russell. Mathematical logic as based on the theory of types. *American Journal of Mathematics*, **30**, 222–62, 1908.

[Russell, 1935–36] B. Russell. The limits of empiricism. *Proceedings of the Aristotelian Society* (new series), **36**, 131–50, 1935–6.

[Salmon, 1970] W. C. Salmon, ed. *Zeno's Paradoxes.* Indianapolis: Hackett, 1970.

[Sazonov, 1995] Y. V. Sazonov. On feasible numbers. In D. Leivant (ed.) *Logic and Computational Complexity.* Lecture Notes in Computer Science, vol. 960. Berlin: Springer-Verlag, pp. 30–51, 1995.

[Scott, 1970] D. S. Scott. Constructive validity. In M. Laudet, D. Lacombe, L. Nolin and M. Schützenberger (eds) *Symposium on Automatic Demonstration (Versailles, December 1968).* Lecture Notes in Mathematics, vol. 125. Berlin: Springer-Verlag, pp. 237–75, 1970.

[Shoenfield, 1977] J. R. Shoenfield. Axioms of set theory. In J. Barwise (ed.) *Handbook of Mathematical Logic.* Amsterdam: North-Holland, pp. 321–44, 1977.

[Skolem, 1922] T. Skolem. Some remarks on axiomatized set theory. In van Heijenoort (1967), pp. 290–301, 1922.

[Smith, 1996] B. Smith. Mereotopology: a theory of parts and boundaries. *Data and Knowledge Engineering*, **20**, no. 3, 287–303, 1996.

[Smith, 1997] B. Smith. Boundaries: an essay in mereotopology. In L.E. Hahn (ed.) *The Philosophy of Roderick M. Chisholm.* Chicago and La Salle, Illinois: Open Court, pp. 533–63, 1997.

[Sochor, 1984] A. Sochor. The alternative set theory and its approach to Cantor's set theory. In H.J. Skala, S. Termini and E. Trillas (eds) *Aspects of Vagueness.* Dordrecht: Reidel, pp. 161–203, 1984.

[Sundholm, 1983] G. Sundholm. Constructions, proofs and the meaning of the logical constants. *Journal of Philosophical Logic*, **12**, no. 2, 151–72, 1983.

[Tait, 1983] W. W. Tait. Against intuitionism: constructive mathematics is part of classical mathematics. *Journal of Philosophical Logic*, **12**, no. 2, 173–95, 1983.

[Tait, 1986] W. W. Tait. Truth and proof: the Platonism of mathematics. *Synthese*, **69**, no. 3, 341–70, 1986.

[Tarski, 1927] A. Tarski. Foundations of the geometry of solids, 1927. Translated by J.H. Woodger. In Tarski, A. (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938.* Oxford: Clarendon Press, pp. 24–9.

[Tharp, 1971] L. H. Tharp. A quasi-intuitionistic set theory. *Journal of Symbolic Logic*, **36**, no. 3, 456–60, 1971.

[Thomson, 1954–55] J. Thomson. Tasks and super-tasks. *Analysis*, **XV**, 1–13. Reprinted in Salmon (1970), pp. 89–102, 1954–5.

[Thomson, 1970] J. Thomson. Comments on Professor Benacerraf's paper. In Salmon (1970), pp. 130–8, 1970.

[Troelstra, 1977] A. S. Troelstra. *Choice Sequences: A Chapter of Intuitionistic Mathematics.* Oxford: Clarendon Press, 1977.

[Troelsta and van Dalen, 1988] A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics: An Introduction.* Amsterdam: North-Holland, 1988.

[van Dalen, 1973] D. van Dalen. Lectures on intuitionism. In A.R.D. Mathias and H. Rogers (eds) *Cambridge Summer School in Mathematical Logic (Cambridge, 1971).* Lecture Notes in Mathematics, vol. 337. Berlin: Springer-Verlag, pp. 1–94, 1973.

[van Heijenoort, 1967] J. van Heijenoort, ed. *From Frege to Gödel: A Source Book in Mathematical Logic 1879–1931.* Cambridge, Massachusetts: Harvard University Press, 1967.

[van Stigt, 1990] W. P. van Stigt. *Brouwer's Intuitionism.* Amsterdam: North-Holland, 1990.

[Varzi, 1996] A. C. Varzi. Parts, wholes, and part-whole relations: the prospects of mereotopology. *Data and Knowledge Engineering*, **20**, no. 3, 259–86, 1996.

[Vieu, 1993] L. Vieu. A logical framework for reasoning about space. *Lecture Notes in Computer Science*, **716**, 25–35, 1993.

[von Neumann, 1925] J. von Neumann. An axiomatization of set theory. Translated by S. Bauer-Mengelberg and D. Føllesdal. In van Heijenoort (1967), pp. 393–413, 1925.

[Vopěnka, 1979] P. Vopěnka. *Mathematics in the Alternative Set Theory.* Leipzig: Teubner, 1979.

[Wang, 1974] H. Wang. *From Mathematics to Philosophy.* London: Routledge and Kegan Paul, 1974.

[Weinberg, 1995] S. Weinberg. *The Quantum Theory of Fields. Vol I: Foundations.* Cambridge: Cambridge University Press, 1995.

[Weinstein, 1983] S. Weinstein. The intended interpretation of intuitionistic logic. *Journal of Philosophical Logic*, **12**, no. 2, 261–70, 1983.

[Wesson, 1991] P. S. Wesson. Olbers's paradox and the spectral intensity of the extragalactic background light. *The Astrophysical Journal*, **367**, 399–406, 1991.

[Weyl, 1921] H. Weyl. On the new foundational crisis of mathematics, 1921. Translated by B. Müller. In [Mancosu, 1998, 86–118].

[Weyl, 1925–27] H. Weyl. The current epistemological situation in mathematics, 1925–7. Translated by B. Müller. In [Mancosu, 1998, 123–42].

[Weyl, 1949] H. Weyl. *Philosophy of Mathematics and Natural Science.* Princeton: Princeton University Press, 1949.

[Whitehead and Russell, 1910] A. N. Whitehead and B. Russell. *Principia Mathematica to *56.* Cambridge: Cambridge University Press, 1910.

[Wright, 1982] C. Wright. Strict finitism. *Synthese*, **51**, no. 2, 203–82, 1982.

[Xia, 1992] Z. Xia. The existence of noncollision singularities in newtonian systems. *Annals of Mathematics*, **135**, no. 3, 411–68, 1992.

[Yablo, 2000] S. Yablo. A reply to new Zeno. *Analysis*, **60**, no. 2, 148–51, 2000.

[Yessenin-Volpin, 1970] A. S. Yessenin-Volpin. The ultra-intuitionistic criticism and the antitraditional program for foundations of mathematics. In A. Kino, J. Myhill and R.E. Vesley (eds) *Intuitionism and Proof Theory.* Proceedings of the summer conference at Buffalo, New York, 1968. Amsterdam: North-Holland, pp. 3–45, 1970.

[Zermelo, 1908] E. Zermelo. Investigations in the foundations of set theory I. Translated by S. Bauer-Mengelberg. In van Heijenoort (1967), pp. 199–215, 1908.

[Zermelo, 1930] E. Zermelo. On boundary numbers and domains of sets: new investigations in the foundations of set theory. Translated by M. Hallett. In Ewald (1996), vol. 2, pp. 1219–33, 1930.

[Zink, 2001] J. Zink. Peirce and the continuum from a philosophical point of view. In P. Schuster, U. Berger and H. Osswald (eds) *Reuniting the Antipodes – Constructive and Nonstandard Views of the Continuum.* Synthese Library, vol. 306. Dordrecht: Kluwer, pp. 303–16, 2001.

LÖWENHEIM-SKOLEM THEOREMS

H.-D. Ebbinghaus

1 INTRODUCTION

In 1915 Leopold Löwenheim proved that every satisfiable first-order sentence has a countable model [Löwenheim, 1915]. His result is regarded as the beginning of model theory. As its proof is rather involved and was outside the focus of the logic of that time, it did not receive instant acknowledgement (cf. [Badesa, 2004]).

The situation changed when Thoralf Skolem generalized Löwenheim's theorem to what is now known as the Löwenheim-Skolem Theorem, namely to countable sets of first-order sentences, providing several simplified proofs [Skolem, 1920; Skolem, 1923; Skolem, 1929] and an unexpected consequence: In his 1923 paper he gives the first formulation of the Zermelo-Fraenkel Axiom System ZF (or ZFC if the Axiom of Choice is included) by formulating Zermelo's Axiom of Separation as demanding that for every *first-order* definable property $\mathcal{E}$ and any set S there exists the set $\{x \in S \mid \mathcal{E}(x)\}$ of the elements x of S having property $\mathcal{E}$, and proceeding similarly with the Axiom of Replacement. He justifies his proposal by its precision, its simplicity, and its sufficient strength. $\mathsf{ZF(C)}$ being a countable system of first-order sentences, he applies the Löwenheim-Skolem Theorem, thus showing that the system — if satisfiable at all — has a countable model, say $\mathcal{M}$. As a model of $\mathsf{ZF(C)}$, $\mathcal{M}$ satisfies all theorems of $\mathsf{ZF(C)}$, in particular the theorem stating that there are uncountable sets and, hence, uncountably many sets. This seeming incompatibility got known as *Skolem's Paradox.* Not clearly understood by the pioneers of axiomatic set theory — Zermelo and Fraenkel — but by the younger representatives such as von Neumann and Gödel, it became the watershed that separated the old generation from the new one which in the 1930s created modern mathematical logic with first-order logic its central part.

Skolem had explicated for the first time and very clearly that first-order logic comes with two dual features: On the one hand it is strong enough to allow a formulation of the whole of set theory and therefore, via modelling mathematical notions in set-theoretic terms, for the whole of mathematics. On the other hand, it lacks the expressive power needed to fix important structures, in particular their cardinality. It was the first aspect, together with Gödel's Completeness Theorem, which led to the rise of first-order logic in the 1930s. Concerning the latter aspect, it turned out that the weakness of first-order logic as illustrated by the Löwenheim-Skolem Theorem (and its generalizations) implied a methodological richness

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

which made the theorem a key fact for first-order model theory. At the end of the 1960s, Lindström [1969] succeeded in proving that first-order logic is optimal in respect to the methodological possibilities it offers: A logic which obeys the Gödel Completeness Theorem and the Löwenheim-Skolem Theorem cannot be strictly stronger than first-order logic.

The Löwenheim-Skolem Theorem can be re-formulated as a special kind of transfer theorem of the type:

If a (set of) sentence(s) has a model of cardinality κ,
then it has a model of cardinality μ.

Given in this way, it leads to a variety of possible Löwenheim-Skolem-type theorems: One may vary κ and μ and the logic. Or one may search for stronger transfer results by asking, given a model $\mathcal{A}$ of cardinality κ, whether the model of cardinalty μ (if it exists) can be chosen as being related to $\mathcal{A}$ in a certain prescribed sense, say, being an extension or a substructure of $\mathcal{A}$. Faced with the fact that the Löwenheim-Skolem Theorem is of major importance for first-order model theory, corresponding theorems may shed light on the model-theoretic possibilities of the logic under consideration.

For a set Φ of sentences of a logic we denote by $\mathsf{spec}(\Phi)$ the *spectrum* of Φ, i. e., the class of cardinalities of its models. Löwenheim-Skolem-type theorems witness certain closure properties of spectra. In this sense the investigation of transfer results for a given logic is related to the problem of describing its spectra.

The present paper is intended to give a picture of what has been obtained with respect to Löwenheim-Skolem phenomena. The contents is as follows:

In §2 we introduce some basic terminology and describe the original result of Löwenheim and Skolem together with immediate extensions to some other logics.

§3 treats stronger transfer theorems for first-order logic and infinitary languages. They are variants of the so-called Löwenheim-Skolem-Tarski Theorem. The section concludes with a glance at the spectrum problem for first-order logic.

§4 investigates special transfers from “big” to “small”. The first part deals with *reflection principles* in set theory; they provide a transfer from “size of a class” to “size of a set”. The second part deals with the *finite model property*, providing a transfer from “infinite” to “finite”.

§5 asks for the validity of transfer theorems for logics which contain operators for fixing cardinalities. The question leads to a more general type of first-order transfer theorems which involve several cardinalities.

After various logics have been considered with respect to the validity of transfer theorems, §6 treats Löwenheim numbers and Hanf numbers as a means to measure the extent to which transfer theorems are valid. It starts by introducing an appropriate framework for a general theory of logical systems.

§7 stays within this framework, describing two results which expose general features of the Löwenheim-Skolem Theorem. The first one treats Lindström-type theorems, thus providing arguments for the essential role of upward and downward transfer for first-order logic. The second one treats a general Löwenheim-Skolem Theorem for logics which are "nearly first-order", a property made precise by set-theoretic conditions which impose some kind of simplicity on the syntax and on the semantics.

The cardinal transfers treated so far concern the case where one starts with an infinite cardinal. Finally, §8 turns to the case where the cardinals involved are finite.

The presentation of the results is accompanied by specific information about the literature. A systematic treatment of first-order Löwenheim-Skolem phenomena can be found, for example, in [Chang and Keisler, 1973] or [Hodges, 1993].

2 THE LÖWENHEIM-SKOLEM THEOREM

2.1 Notational conventions

The set of natural numbers $0, 1, 2, \dots$ is denoted by ω.

Vocabularies τ consist of *relation symbols* $R, \dots$, *function symbols* $f, \dots$, and/or *constants* $c, \dots$. Each $R \in \tau$ has an arity $\mathsf{ar}(R) \geq 1$ and each $f \in \tau$ an arity $\mathsf{ar}(f) \geq 1$. A vocabulary τ is *relational* if it contains only relation symbols.

To define *first-order logic* FO, we fix a denumerable set $V = \{x_i \mid i \in \omega\}$ of first-order variables.

First-order τ-*formulas* are defined as usual and may contain the identity symbol. $\mathsf{FO}[\tau]$ is the set of (first-order) τ-formulas. A formula without a free variable is called a *sentence.* By $\varphi(x_1, \dots, x_n)$ or, for short, $\varphi(\overline{x})$, we denote formulas φ where the variables free in φ occur among $x_1, \dots x_n$. The abbreviation $\overline{x}$ always refers to tuples of appropriate length.

The quantifier rank of a formula is the maximal number of nested quantifiers occurring in it.

For a sentence φ and a unary relation symbol U, φ^U denotes the *relativization* of φ to U. It arises from φ by relativizing all quantifiers to U. Intuitively, φ^U says that φ holds on U.

With $\tau = \{R, \dots, f, \dots, c, \dots\}$, a τ-structure $\mathcal{A}$ is of the form

$$\mathcal{A} \;=\; (A, R^{\mathcal{A}}, \dots, f^{\mathcal{A}}, \dots, c^{\mathcal{A}}, \dots)$$

where $A \neq \emptyset$ is the *domain* of $\mathcal{A}$, and for $R \in \tau$, $f \in \tau$, and $c \in \tau$ we have $R_{\mathcal{A}} \subseteq A^{\mathsf{ar}(R)}$, $f^{\mathcal{A}} : A^{\mathsf{ar}(f)} \to A$, and $c^{\mathcal{A}} \in A$, respectively.

A τ-structure $\mathcal{A}$ is a *substructure* of the τ-structure $\mathcal{B}$, in terms: $\mathcal{A} \subseteq \mathcal{B}$, if $A \subseteq B$ and $\mathcal{A}$ and $\mathcal{B}$ coincide on A. In this case, A is τ-*closed* in $\mathcal{B}$, i. e., $c \in A$ for $c \in \tau$ and A is closed under $f^{\mathcal{B}}$ for $f \in \tau$.

If σ is a vocabulary with $\sigma \supseteq \tau$, $\mathcal{A}$ a τ-structure, and $\mathcal{B}$ a σ-structure, we write

$$\mathcal{A} = \mathcal{B} \restriction \tau$$

if $A = B$ and $\mathcal{A}$ and $\mathcal{B}$ coincide on τ. We then call $\mathcal{A}$ a *reduct* (the τ-*reduct*) of $\mathcal{B}$ and $\mathcal{B}$ an *expansion* of $\mathcal{A}$.

For a cardinal κ, $\mathcal{A}$ has power κ if the cardinality $|A|$ of A equals κ.

Given a τ-structure $\mathcal{A}$, an $\mathcal{A}$-assignment α of the set V of variables is a mapping from V into the domain A of $\mathcal{A}$. For $\mathcal{A}$, α, and $\varphi \in \mathsf{FO}[\tau]$, say, $\varphi = \varphi(\overline{x})$, we write

$$(\mathcal{A}, \alpha) \models \varphi \quad \text{or} \quad \mathcal{A} \models \varphi(\alpha(\overline{x}))$$

if $(\mathcal{A}, \alpha)$ satisfies φ or is a model of φ in the usual sense (where the variables $\overline{x}$ are interpreted by $\alpha(\overline{x})$).

Given $\overline{a} \in A$ (short for $\overline{a} \in A^{\text{length of } \overline{a}}$),

$$\mathcal{A} \models \varphi(\overline{a})$$

means that $(\mathcal{A}, \alpha) \models \varphi$ for any $\mathcal{A}$-assignment α with $\alpha(\overline{x}) = \overline{a}$.

In case φ is a sentence, we simply write $\mathcal{A} \models \varphi$.

By definition, $(\mathcal{A}, \alpha)$ is a model of the set Φ of first-order τ-formulas, in terms: $(\mathcal{A}, \alpha) \models \Phi$, if $(\mathcal{A}, \alpha)$ is a model of all $\varphi \in \Phi$.

A τ-formula or a set of τ-formulas is *satisfiable* if it has a model.

A set is *countable* if it is finite or countably infinite. If κ is a cardinal, κ^+ denotes the cardinal successor of κ, i. e., the smallest cardinal $> \kappa$.

2.2 The theorem and its proof

The proof of the Löwenheim-Skolem Theorem sketched below follows a simple idea: Given a large model $\mathcal{A}$ of a set Φ of first-order sentences, one can construct a small model $\mathcal{B}$ of Φ by choosing a small subset of the domain A of $\mathcal{A}$ and closing it under examples for first-order existential assertions true in $\mathcal{A}$. If B is the closure, one can define $\mathcal{B}$ as the substructure of $\mathcal{A}$ with domain B.

The Löwenheim-Skolem Theorem. If Φ is a countable set of first-order τ-sentences which is satisfiable (say, by the τ-structure $\mathcal{A}$) then Φ has a countable model (which can be chosen as a substructure of $\mathcal{A}$).

Proof. We may assume that τ is countable, too. To illustrate the main idea, we first treat the special case where $\tau = \{R\}$ with binary R and $\Phi = \{\forall x\, \exists y\, Rxy\}$. Let

$$\mathcal{A} \models \forall x\, \exists y\, Rxy.$$

In order to get a countable substructure $\mathcal{B} \subseteq \mathcal{A}$ with $\mathcal{B} \models \forall x\, \exists y\, Rxy$ we start with a nonempty countable subset B_0 of A, say

$$B_0 = \{a_0\},$$

where $a_0 \in A$, and define an ascending chain $B_0 \subseteq B_1 \subseteq B_2 \ldots$ of countable subsets of A by choosing, for any $i \in \omega$, B_{i+1} above B_i in such a way that for all $b \in B_i$, the set B_{i+1} contains some b' with $(b, b') \in R^{\mathcal{A}}$. Let $B := \bigcup_{i \in \omega} B_i$. As τ is relational, B is τ-closed in $\mathcal{A}$ and, hence, the domain of a countable substructure $\mathcal{B}$ of $\mathcal{A}$. By construction, $\mathcal{B} \models \forall x\, \exists y\, Rxy$.

The definition of the chain $B_0 \subseteq B_1 \subseteq B_2 \ldots$ can be given in a slightly different form by choosing a so-called *Skolem function* for $\forall x\, \exists y\, Rxy$ in $\mathcal{A}$, namely a function $h : A \to A$ such that for all $a \in A$,

$$h(a) := \text{ some } b \text{ such that } \mathcal{A} \models Rab.$$

Then one can define B_{i+1} as the union of B_i and its image under h, and B is the closure of B_0 in $\mathcal{A}$ under h.

If Φ consists of sentences of the form $\forall \overline{x}\, \exists \overline{y}\, \varphi(\overline{x}, \overline{y})$ with quantifier-free φ (so-called Π_2-sentences or sentences in *Skolem normal form*), we can argue as above. In his 1920 proof Skolem reduces the general case to this special case. Here we sketch a somewhat different way which will pay later.

Let τ be countable, $\Phi \subseteq \mathsf{FO}[\tau]$, and $\mathcal{A} \models \Phi$. Let Σ be the set of all formulas of the form $\exists y\, \psi(\overline{x}, y)$ where ψ is an arbitrary τ-formula. Σ is countable. For

$$\exists y\, \psi(\overline{x}, y) =: \chi \in \Sigma$$

we define a Skolem function $h_\chi : A^{\text{length of } \overline{x}} \to A$, using a fixed element a_0 of A, as follows:

$$h_\chi(\overline{a}) \ := \ \begin{cases} a_0, \text{ if } \mathcal{A} \models \neg \exists y\, \psi(\overline{a}, y) \\ \text{some } b \text{ such that } \mathcal{A} \models \psi(\overline{a}, b), \text{ else.} \end{cases}$$

We construct the chain $B_0 \subseteq B_1 \subseteq B_2 \subseteq \ldots \subseteq A$ with B_0 as above by setting

$$B_{i+1} \ := \ B_i \cup \{h_\chi(\overline{a}) \mid \chi \in \Sigma \text{ and } \overline{a} \in B_i\}$$

for all $i \in \omega$ and $B := \bigcup_{i \in \omega} B_i$. It is easy to show that B is τ-closed and, hence, induces a (countable) substructure $\mathcal{B}$ of $\mathcal{A}$ with domain B. By induction on the composition of formulas one can prove that for all $\varphi(\overline{x})$ and for all $\overline{b} \in B$,

$$\mathcal{A} \models \varphi(\overline{b}) \ \text{ iff } \ \mathcal{B} \models \varphi(\overline{b}).$$

As $\mathcal{A} \models \Phi$, we have $\mathcal{B} \models \Phi$. ■

Our argumentation (as also that in [Skolem, 1920]) uses the Axiom of Choice, namely in the definition of the h_χ. The argumentation in [Skolem, 1923] avoids the Axiom of Choice by constructing the chain $B_0 \subseteq B_1 \subseteq B_2 \subseteq \ldots$ in the set of natural numbers, making use of their natural well-ordering. Hence, when starting with a model $\mathcal{A}$ of Φ, it leads to a countable model $\mathcal{B}$ whose domain is a set of natural numbers. We will not pay attention to the Axiom of Choice, but will tacitly argue in ZFC.

As mentioned in the introduction, the Löwenheim-Skolem Theorem illustrates a weakness of first-order logic, at the same time playing an essential role in first-order model theory. To give an illustration of the latter aspect, we show that the theory $T \subseteq \mathsf{FO}[\{<\}]$ of dense open orderings, i. e., of orderings which are dense and do not have a minimal or a maximal element (like the ordering of the rationals), is complete in the sense that for any sentence $\varphi \in \mathsf{FO}[\{<\}]$, either $\varphi \in T$

or $\neg\varphi \in T$. For *reductio ad absurdum* assume that for some sentence φ neither $\varphi \in T$ nor $\neg\varphi \in T$. Then both $T \cup \{\varphi\}$ and $T \cup \{\neg\varphi\}$ are satisfiable. By the Löwenheim-Skolem Theorem they have countable models $\mathcal{A}$ and $\mathcal{B}$, respectively. By a theorem of Cantor, any two countable dense open orderings are isomorphic. Hence, $\mathcal{A} \models \varphi$ iff $\mathcal{B} \models \varphi$, a contradiction.

2.3 Some immediate extensions

The proof of the Löwenheim-Skolem Theorem given above can immediately be extended to some more powerful logics. We give two examples.

Logic with the Infinity Quantifier. Let Q be a new unary quantifier symbol and extend the rules for forming first-order formulas by the rule allowing the transition from φ to $Qx\,\varphi$ for any variable x. Extend the satisfaction relation by

$$\begin{gathered}\mathcal{A} \models Qx\,\varphi(\overline{a}, x) \text{ iff}\\ \text{there are infinitely many } b \in A \text{ such that } \mathcal{A} \models \varphi(\overline{a}, b).\end{gathered}$$

The resulting logic $\mathsf{FO}(Q)$ is stronger than first-order logic. For example, there is no $\Phi \subseteq \mathsf{FO}[\emptyset]$ whose models are just the finite sets, whereas the sentence $\neg Qx\, x = x$ of $\mathsf{FO}(Q)$ has this property. To prove the Löwenheim-Skolem Theorem for $\mathsf{FO}(Q)$, one proceeds as in the proof above, letting Σ contain also the corresponding formulas of the form $Qy\,\psi(\overline{x}, y)$ and adding to B_{i+1} countably many b with $\mathcal{A} \models \psi(\overline{a}, b)$ in case $\overline{a} \in B_i$ and $\mathcal{A} \models Qy\,\psi(\overline{a}, b)$.

Weak Second-Order Logic. Second-order logic SO does not satisfy the analogue of the Löwenheim-Skolem Theorem. For instance, the ordered field of the reals is up to isomorphism the only model of the (finite) set Φ consisting of the axioms for ordered fields together with the Completeness Axiom saying that *every subset* of the universe which is nonempty and bounded to the left, has an infimum. So Φ has only models of power $2^{\aleph_0}$.

Things change if we consider *weak second order logic* WSO. In WSO no function variables are allowed, and the relation variables run over *finite* relations. WSO satisfies the analogue of the Löwenheim-Skolem Theorem. The proof for FO extends to WSO in a way similar to the extension from FO to $\mathsf{FO}(Q)$, making use of the simple fact that a finite relation comprises only finitely many elements.

3 LÖWENHEIM-SKOLEM-TARSKI THEOREMS

3.1 The downward Löwenheim-Skolem-Tarski theorem

We start with a definition. Let $\mathcal{A}$ and $\mathcal{B}$ be τ-structures. $\mathcal{B}$ is an *elementary substructure* of $\mathcal{A}$ (and $\mathcal{A}$ an *elementary extension* of $\mathcal{B}$), in terms: $\mathcal{B} \preceq \mathcal{A}$, if $\mathcal{B} \subseteq \mathcal{A}$ and for all $\varphi(\overline{x}) \in \mathsf{FO}[\tau]$ and $\overline{b} \in B$,

$$\mathcal{B} \models \varphi(\overline{b}) \quad \text{iff} \quad \mathcal{A} \models \varphi(\overline{b}).$$

If $\mathcal{B} \preceq \mathcal{A}$, then $\mathcal{A}$ and $\mathcal{B}$ satisfy the same first-order τ-sentences even if parameters from B are allowed. In general, the elementary substructure relation is stronger than the substructure relation. For example, the field $\mathcal{Q}$ of the rational numbers is a substructure of the field $\mathcal{R}$ of the real numbers, but not an elementary one, as 2 is a square in $\mathcal{R}$, but not in $\mathcal{Q}$.

With a straightforward generalization to higher cardinals the proof of the Löwenheim-Skolem Theorem given above yields the following stronger form which goes back to Tarski [Tarski and Vaught, 1957]:

THEOREM. For any τ-structure $\mathcal{A}$ and any infinite cardinal λ, if $|\mathsf{FO}[\tau]| \leq \lambda \leq |A|$, then there is a τ-structure $\mathcal{B}$ with

$$\mathcal{B} \preceq \mathcal{A} \text{ and } |B| = \lambda.$$

$\mathcal{B}$ can be chosen such that $X \subseteq B$ for a given subset X of A of cardinality $\leq \lambda$.

Moreover, there is an extension to certain infinitary logics. For a given infinite cardinal κ, let the logic $\mathsf{L}_{\kappa^+\omega}$ be defined like first-order logic, but allowing conjunctions and disjunctions over sets of formulas of cardinality $\leq \kappa$ as far as the resulting formulas contain only finitely many free variables. For every formula φ of $\mathsf{L}_{\kappa^+\omega}$ there is a vocabulary τ with $|\tau| \leq \kappa$ such that $\varphi \in \mathsf{L}_{\kappa^+\omega}[\tau]$; moreover, φ has at most κ many subformulas.

As usual, we write $\mathsf{L}_{\omega_1\omega}$ for $\mathsf{L}_{\aleph_1\omega}$.

A strict analogue of the Löwenheim-Skolem-Tarski Theorem fails already in the countable case. For example, any $\mathsf{L}_{\omega_1\omega}$-elementary substructure $\mathcal{B}$ of the ordered field $\mathcal{R}$ of the reals (as a $\{+, \cdot, 0, 1, <\}$-structure) contains all reals since each real x can be described by countably many inequalities of the form $x < \rho$ or $\rho < x$ with rational numbers ρ.[1]

There is, however, a weaker form: Given, say, a relational τ, a sentence φ of $\mathsf{L}_{\kappa^+\omega}[\tau]$, and a model $\mathcal{A}$ of φ, one can proceed as in the proof of the Löwenheim-Skolem Theorem, taking as Σ the set of all formulas $\exists y\, \psi(\overline{x}, y)$ of $\mathsf{L}_{\kappa^+\omega}[\tau]$ where ψ is a subformula of φ or the negation of such a formula. Σ has cardinality $\leq \kappa$. Concentrating on the subformulas of φ and their negations, one thus gets the relational version of

Downward Löwenheim-Skolem Theorem for Infinitary Logic. Let κ be infinite, $|\tau| \leq \kappa$, and φ a sentence in $\mathsf{L}_{\kappa^+\omega}[\tau]$. If $\mathcal{A}$ is a τ-structure with $\mathcal{A} \models \varphi$ and $\kappa \leq \lambda \leq |A|$, then there is a τ-structure $\mathcal{B}$ with

$$\mathcal{B} \subseteq \mathcal{A},\ \mathcal{B} \models \varphi, \text{ and } |B| = \lambda.$$

If τ is not relational, we enlarge Σ by the formulas $\exists y\, f(\overline{x}) = y$ and $\exists y\, c = y$ for $f, c \in \tau$. It is here that the assumption $|\tau| \leq \kappa$ is needed.

To give an application, we consider simple groups, i. e., groups having no nontrivial normal subgroups. The class of simple groups can be axiomatized by a sentence of $\mathsf{L}_{\omega_1\omega}$. By the preceding theorem, each simple group $\mathcal{G}$ has a countable simple subgroup.

---

[1] For, say, $\rho = 2/3$, $x < \rho$ can be expressed as $x + x + x < 1 + 1$.

3.2 The upward Löwenheim-Skolem-Tarski theorem

The transfer theorems given so far deal with downward transfers. What about the other direction? In the first-order case it can be proved by making use of the compactness theorem. In §§6 and 7 we shall see that upward transfer and compactness are related in various ways.

We start with a simple case: If Φ is a set of first-order τ-sentences and $\mathcal{A}$ an infinite model of Φ, then by the compactness theorem for first-order logic, given a set C of new constants of power $\geq \lambda$, the set

$$\Phi \cup \{\neg c = c' \mid c, c' \in C,\, c \neq c'\}$$

has a model, say $\mathcal{B}$. Obviously, $|B| \geq \lambda$. We get more if we work not with Φ, but with the so-called *elementary diagram* $\mathsf{ED}(\mathcal{A})$ of the model $\mathcal{A}$ of Φ. With new constants $\boldsymbol{a}$ for $a \in A$ we set

$$\mathsf{ED}(\mathcal{A}) \;:=\; \{\varphi(\overline{\boldsymbol{a}}) \mid \varphi(\overline{x}) \in \mathsf{FO}[\tau],\, \overline{a} \in A,\, \mathcal{A} \models \varphi(\overline{a})\}.$$

(Note that $\Phi \subseteq \mathsf{ED}(\mathcal{A})$.) Then the corresponding model $(\mathcal{B}, (\boldsymbol{a}^{\mathcal{B}})_{a \in A})$ with $|B| \geq \lambda$ can be chosen such that $\boldsymbol{a}^{\mathcal{B}} = a$ for $a \in A$, i. e., such that $\mathcal{A} \preceq \mathcal{B}$. If $|A| \leq \lambda$ and $\mathsf{FO}[\tau] \leq \lambda$, we apply the Downward Löwenheim-Skolem-Tarski Theorem to $\mathcal{B}$ and $X := A$ to get $\mathcal{D}$ with $\mathcal{A} \preceq \mathcal{D} \preceq \mathcal{B}$ and $|D| = \lambda$. Hence, we have proved the following theorem from [Tarski and Vaught, 1957]:

Upward Löwenheim-Skolem-Tarski Theorem. Let $\mathcal{A}$ be an infinite τ-structure and $\lambda \geq \mathsf{max}\{|A|, |\mathsf{FO}[\tau]|\}$. Then there is a τ-structure $\mathcal{B}$ with

$$\mathcal{A} \preceq \mathcal{B} \text{ and } |B| = \lambda. \qquad \dashv$$

The extensions of first-order logic considered so far do not allow a corresponding upward transfer: For any infinite κ, the logic $\mathsf{L}_{\kappa^+\omega}$ allows to characterize the natural ordering of ω up to isomorphism, namely by the axioms for orderings with a first but not a last element together with the sentence

$$\forall x \bigvee \{\text{“there are at most } n \text{ elements} < x\text{”} \mid n = 0, 1, 2, \ldots\}.$$

Similarly one can argue with $\mathsf{FO}(Q)$ and weak second-order logic WSO.

The Löwenheim-Skolem-Tarski Theorems can be regarded as strengthenings of the simple transfer theorem addressed in the introduction, requiring that the models in question are linked by the elementary substructure relation. There is a variety of other strengthenings. For example, the models may have additional properties such as being saturated, admitting few types, *etc.* Many theorems of first-order model theory which show the existence of models with special properties, can be interpreted in this way. For later reference, Subsection 3. 4 below contains an important example: models with indiscernibles.

3.3 Spectra

Let Φ be a set of first-order τ-sentences with an infinite model. According to the Löwenheim-Skolem-Tarski Theorems, Φ has models of all cardinalities $\geq |\mathsf{FO}[\tau]|$, i. e., the *spectrum* of Φ,

$$\mathsf{spec}(\Phi) \ := \ \{|A| \mid \mathcal{A} \models \Phi\},$$

the class of cardinalities of models of Φ, contains all cardinals $\geq |\mathsf{FO}[\tau]|$. What about the cardinalities of models of Φ in the interval

$$[\omega, |\mathsf{FO}[\tau]|\rangle \ := \ \{\nu \mid \omega \leq \nu < |\mathsf{FO}[\tau]|\}?$$

A first answer is given by the following result of Keisler (cf. [Chang and Keisler, 1973], Cor. 6. 5. 12):

THEOREM. For all $\mu \geq \omega$ and $\nu \geq \mu^\omega$, if μ is less than the first (uncountable) measurable cardinal and $\mu \in \mathsf{spec}(\Phi)$ then $\nu \in \mathsf{spec}(\Phi)$.

The special case $\nu = \mu^\omega$ is a consequence of a weaker form of the theorem saying that $\mathsf{spec}(\Phi)$ is closed under exponentiations in the infinite. The *proof* of this form uses ultrapower techniques: Let $\mathcal{A} \models \Phi$ with $|A| = \mu \geq \omega$, and let $\nu \geq \omega$. Moreover, let D be a regular ultrafilter over ν. Then the ultrapower $\prod_{\iota\in\nu} \mathcal{A}/D$ is a model of Φ of cardinality μ^ν.

Results concerning the possible gap between μ and μ^ω in the theorem would have to be of a special kind. A result of Mekler [1977] shows that for $\mu \in \mathsf{spec}(\Phi)$ the set $\mathsf{spec}(\Phi) \cap [\mu, \mu^\omega\rangle$ has no regular pattern: Given μ less than the first measurable cardinal such that

$$\omega \leq \mu < \mu^\omega,$$

and given $S \subseteq [\mu, \mu^\omega\rangle$, there are τ and a set Φ of first-order τ-sentences such that

$$\mathsf{spec}(\Phi) \cap [0, \mu^\omega\rangle = S.$$

3.4 Appendix: models with indiscernibles

Indiscernibles provide a powerful tool in model theory and will be of importance for several topics treated below. In order to give an impression, we describe a basic variant in more detail.

Let $\mathcal{A}$ be a τ-structure and $(B, <)$ an ordering with $B \subseteq A$. $(B, <)$ is an *ordered set of indiscernibles* in $\mathcal{A}$ if, for all $\varphi(\overline{x}) \in \mathsf{FO}[\tau]$ and all $<$-ascending tuples $\overline{a}, \overline{b} \in B$,[2]

$$\mathcal{A} \models \varphi(\overline{a}) \ \text{ iff } \ \mathcal{A} \models \varphi(\overline{b}).$$

The main result, going back to Ehrenfeucht and Mostowski [1956], says:

PROPOSITION. If $\Phi \subseteq \mathsf{FO}[\tau]$ is a set of sentences with an infinite model and $(C, <)$ is an ordering, then Φ has a model which has $(C, <)$ as an ordered set of indiscernibles. If $|C| \geq |\mathsf{FO}[\tau]|$, then there is such a model which has cardinality $|C|$.

---

[2] Equivalently, one can consider $<$-isomorphic tuples.

Proof. Choosing new constants $\boldsymbol{c}$ for $c \in C$, it suffices to show that the set

$$\Theta := \Phi \cup \big\{\boldsymbol{c} \neq \boldsymbol{d} \mid c, d \in C,\, c \neq d\big\}$$
$$\cup \big\{\varphi(\overline{\boldsymbol{c}}) \leftrightarrow \varphi(\overline{\boldsymbol{d}}) \mid \varphi(\overline{x}) \in \mathsf{FO}[\tau] \text{ and } \overline{c}, \overline{d} \text{ <-ascending tuples over } C\big\}$$

has a model. W.l.o.g. one can assume that $(C, <)$ is dense and open. By the compactness theorem the satisfiability of Θ can be reduced to the satisfiability of all sets of the form $\Phi \cup \Psi$ with Ψ a *finite* set of inequalities $\boldsymbol{c} \neq \boldsymbol{d}$ and of equivalences

$$\varphi(\overline{\boldsymbol{c}}) \;\leftrightarrow\; \varphi(\overline{\boldsymbol{d}}),$$

where $\overline{c}, \overline{d} \in C$ are $<$-ascending and for the finitely many $\varphi_1(\overline{x}), \ldots, \varphi_k(\overline{x})$ involved, the tuples $\overline{x}$ are of the same length $n(\Psi)$.

Let $\mathcal{A}$ be an infinite model of Φ and $\prec$ an arbitrary ordering relation on A. Given Ψ as above, define an equivalence relation $\sim$ on the set of $\prec$-ascending $n(\Psi)$-tuples of A by setting

$$\overline{a} \sim \overline{b} \quad \text{:iff} \quad \big(\mathcal{A} \models \varphi_i(\overline{a}) \;\text{ iff } \mathcal{A} \models \varphi_i(\overline{b})\big) \text{ for } i = 1, \ldots, k.$$

As $\sim$ has only finitely many equivalence classes, Ramsey's Theorem yields that A contains an infinite subset B such that all $\prec$-ascending $n(\Psi)$-tuples over B belong to the same equivalence class. Hence, $\Phi \cup \Psi$ has a model which results from $\mathcal{A}$ by interpreting the constants $\boldsymbol{c}$ in Ψ by suitable $b \in B$. ■

4 FROM "VERY LARGE" TO "VERY SMALL"

Basically, the Löwenheim-Skolem Theorem deals with transfer from infinite cardinalities to smaller infinite cardinalities. In this section we shortly discuss transfer from "very large" to "very small", exemplifying the contrast by (1) "very large" = "size of a class" and "very small" = "size of a set" and (2) "very large" = "infinite" and "very small" = "finite".

4.1 The reflection theorem

We consider first-order set theory based on the Zermelo-Fraenkel Axiom System ZF. The levels V_α of the von Neumann Hierarchy are defined by iterating the power set operation along the ordinals:

$$V_0 = \emptyset$$
$$\text{for all ordinals } \alpha,\; V_{a+1} = \text{ powerset of } V_\alpha$$
$$\text{for all limit ordinals } \delta,\; V_\delta = \bigcup\{V_\beta \mid \beta < \delta\}.$$

The hierarchy is *cumulative* in the sense that $V_\alpha \subseteq V_\beta$ for $\alpha \leq \beta$. Hence, the union of a set of levels is a level. The union of all V_α is the whole universe.

Let $\varphi(\overline{x})$ be a first-order formula of set theory. Regarding the universe as a big $\{\in\}$-structure, one can imitate the proof of the Löwenheim-Skolem Theorem, thereby concentrating on the subformulas of φ and their negations (as in the proof of the infinitary version in Subsection 3.1.) and starting with an arbitrary level V_γ. Instead of adding single examples for a valid statement $\exists y\, \psi(\overline{a}, y)$, one adds all elements of the lowest level V_β containing an example, in this way avoiding use of the Axiom of Choice. Let W be the union of all levels V_β involved in the construction. Then W is a level which behaves as an elementary subset of the universe (with respect to the $\in$-relation) for the subformulas of φ. By writing

$$\text{“}V_\beta \text{ reflects } \varphi(\overline{x})\text{”} \quad \text{for} \quad \text{“}\forall\, \overline{x} \in V_\beta \left(\varphi(\overline{x})^{V_\beta} \leftrightarrow \varphi(\overline{x})\right)\text{”},$$

we thus have:

Reflection Theorem. Let $\varphi(\overline{x})$ be a first-order formula of set theory. Then ZF proves that for all γ there exists an $\alpha \geq \gamma$ such that V_α reflects $\varphi(\overline{x})$. In particular, if φ is a sentence which is provable in ZF, then ZF proves that φ holds in some level of the von Neumann Hierarchy.

The theorem goes back to Montague [1961] and Levy [1960]. Levy showed in addition that reflection is equivalent to the Axiom of Replacement together with the Axiom of Infinity.

By Gödel's Second Incompleteness Theorem one cannot prove that the universe (together with the $\in$-relation) has an elementary substructure V_α. So the restriction to single formulas in the Reflection Theorem is essential. On the other hand, stronger forms of reflection may serve as stronger axioms of infinity (cf., e.g., [Kanamori, 1994]).

4.2 The finite model property

Let Φ be a class of formulas of a logic L. Φ has the *finite model property* if any satisfiable $\varphi \in \Phi$ has a finite model.

In the case of first-order logic, the finite model property for a recursive Φ implies the decidability of the satisfiability problem for Φ. If, say, τ is finite and Φ a recursive set of first-order τ-sentences with the finite model property, the satisfiable sentences of Φ are recursively enumerable as one can enumerate the sentences in Φ and all finite τ-structures (up to isomorphism) and effectively check the satisfaction relation; the unsatisfiable sentences of Φ are recursively enumerable because their negations are tautologies and, hence, recursively enumerable by the Gödel Completeness Theorem.

As a rule, one gets complexity results, too. To give some examples, let

- τ be finite and relational and Φ the set of universal first-order τ-sentences, i.e., of τ-sentences $\forall\, \overline{x}\, \varphi$ with quantifier-free φ;
- $\tau = \{P_1, \ldots, P_n\}$ with unary relation symbols P_i and Φ the set of first-order τ-sentences.

In order to show the finite model property of Φ, observe for the first example that the model class of a universal first-order sentence is closed under substructures. Concerning the second example, given a τ-sentence φ, let $\mathcal{A}$ and $\mathcal{B}$ be τ-structures which are *φ-similar* in the sense that for any conjunction $(\neg)P_1x \wedge \ldots \wedge (\neg)P_nx$ the number of elements of $\mathcal{A}$ and the number of elements of $\mathcal{B}$ satisfying it are equal or both greater than the quantifier rank $\mathsf{qr}(\varphi)$ of φ. Then one can prove that $\mathcal{A} \models \varphi$ iff $\mathcal{B} \models \varphi$. For any $\mathcal{A}$ there is a finite $\mathcal{B}$ of size $\leq 2^n \cdot (1 + \mathsf{qr}(\varphi))$ such that $\mathcal{A}$ and $\mathcal{B}$ are φ-similar.

Hence, in both cases there exists a recursive function f defined on Φ with values in ω such that any satisfiable $\varphi \in \Phi$ has a model of cardinality $\leq f(\varphi)$. Thus, by checking "$\mathcal{A} \models \varphi$?" for all τ-structures with domain $\subseteq \{0, \ldots, f(\varphi) - 1\}$, one can decide the satisfiability of φ and also evaluate the complexity of this decision procedure, thus getting an upper bound for the complexity of the satisfiability problem.

The decidability of the satisfiability problem for first-order sentences containing only unary relation symbols goes back to [Löwenheim, 1915]. For further results see [Börger *et al.*, 1997]. We conclude with an important example of Mortimer [1975], the complexity bound stemming from [Grädel *et al.*, 1997]:

THEOREM. First-order logic restricted to sentences of a relational vocabulary which use at most two different variables, has the finite model property. More exactly: Any such sentence φ which is satisfiable, has a model of a cardinality which is exponentially bounded in the length of φ.

Two-variable first-order logic is strong enough to admit an effective embedding of various modal logics. Hence, also these logics obey an effective version of the finite model property and have a decidable satisfiability problem (cf. [Gabbay, 1971] and, for temporal logics, also [Etessami *et al.*, 2002]).

5 BUILDING CARDINALITIES INTO THE LOGIC

In order to overcome the weakness of first-order logic apparent in the Löwenheim-Skolem-type theorems, extensions of first-order logic with built-in possibilities for fixing cardinalities have been introduced. We shall consider here the most basic type of such logics. They will lead to theorems on the simultaneous transfer for several cardinals.

5.1 Logic with the quantifier "There are $\aleph_\alpha$ Many"

In Subsection 2. 2 we have described the logic $\mathsf{FO}(Q)$ which arises from first-order logic by adjoining a new unary quantifier Q interpreted by "there are infinitely many". In the following we write "Q_0" instead of "Q", thereby indicating the intended meaning "there are at least $\aleph_0$ many". Generalization from $\alpha = 0$ to arbitrary ordinals α results in the logics $\mathsf{FO}(Q_\alpha)$, where Q_α now says that there are at least $\aleph_\alpha$ many.

The logics $\mathsf{FO}(Q_\alpha)$ go back to Mostowski [1957]. They were first investigated by Fuhrken [1964; 1965] and Vaught [1964]. For information about them and the variety of further logics with cardinality constructs, cf. [Ebbinghaus, 1985; Kaufmann, 1985; Mundici, 1985]. We mainly treat the logic $\mathsf{FO}(Q_1)$ here.

$\mathsf{FO}(Q_1)$ satisfies the compactness theorem for countable sets of formulas and a completeness theorem for countable sets of premises. An elegant calculus was given by Keisler [1970]. On the other hand, compactness fails already for sets of sentences of cardinality $\aleph_1$ as exemplified by the set

$$\{\neg Q_1 x\, x = x\} \cup \{\neg c_\alpha = c_\beta \mid \alpha < \beta < \aleph_1\}.$$

The exact analogue of the Löwenheim-Skolem Theorem fails, too, as exemplified by the sentence $Q_1 x\, x = x$. Even more, $\mathsf{FO}(Q_1)$ is able to fix the cardinality $\aleph_1$, say, by a $\{<\}$-sentence expressing that (universe, $<$) is an $\aleph_1$-*like ordering*, i. e., an uncountable ordering such that every strict initial segment is countable. However, there is the following weaker form which ensures transfer down to $\aleph_1$:

THEOREM. Let τ be of cardinality $\leq \aleph_1$ and $\mathcal{A}$ an uncountable τ-structure. Then for any κ with $\aleph_1 \leq \kappa \leq |A|$, $\mathcal{A}$ has an $\mathsf{FO}(Q_1)$-elementary substructure of cardinality κ.

Proof. It varies that for $\mathsf{FO}(Q)$ in Subsection 2. 3 by starting with a subset of cardinality κ and choosing $\aleph_1$ many examples for true statements of the form $Q_1 y\, \varphi(\overline{a}, y)$. ■

Obviously the theorem and its proof can be generalized to all $\mathsf{FO}(Q_\alpha)$.

The analogue of the Upward Löwenheim-Skolem-Tarski Theorem fails in a strong form. For given κ we define the corresponding *Beth operation* recursively by

$$\begin{aligned} \beth_0(\kappa) &= \kappa \\ \text{for all ordinals } \alpha,\ \beth_{\alpha+1}(\kappa) &= 2^{\beth_\alpha(\kappa)} \\ \text{for all limit ordinals } \delta,\ \beth_\delta(\kappa) &= \mathsf{sup}\{\beth_\beta(\kappa) \mid \beta < \delta\}. \end{aligned}$$

Then we have:

PROPOSITION. For every $n \in \omega$ there is a vocabulary τ_n and a sentence φ_n of $\mathsf{FO}(Q_1)[\tau_n]$ which has a model of cardinality $\beth_n(\aleph_0)$, but no larger model.

Proof. We set $\tau_0 := \emptyset$ and $\varphi_0 := \neg Qx\, x = x$. Suppose that τ_n and φ_n are already defined and satisfy the claim. Using a new unary relation symbol U and a new binary relation symbol E, we set

$$\tau_{n+1} := \tau_n \cup \{U, E\}$$

and let φ_{n+1} say that φ_n holds in the substructure with domain U and that the elements of the universe code subsets of U, Exy meaning that x belongs to the subset of U coded by y, i. e., we set

$$\varphi_{n+1} := \varphi_n^U \wedge \forall xy \left(Exy \rightarrow Ux\right) \wedge \forall yy' \left(\forall x \left(Exy \leftrightarrow Exy'\right) \rightarrow y = y'\right).$$

Then in any model $\mathcal{A}$ of φ_{n+1} we have

$$|U^{\mathcal{A}}| \leq \beth_n(\aleph_0) \quad \text{and} \quad |A| \leq 2^{|U^{\mathcal{A}}|},$$

and, hence,

$$|A| \leq 2^{\beth_n(\aleph_0)} = \beth_{n+1}(\aleph_0).$$

As φ_n has a model of cardinality $\beth_n(\aleph_0)$, the sentence φ_{n+1} has a model of cardinality $\beth_{n+1}(\aleph_0)$. ■

5.2 Two cardinal theorems

Given α and β, one can ask whether $\mathsf{FO}(Q_\alpha)$ and $\mathsf{FO}(Q_\beta)$ have the same valid sentences, or, more general, to what extent $\mathsf{FO}(Q_\alpha)$ and $\mathsf{FO}(Q_\beta)$ behave in the same way. As for any α the quantifier Q_α acts trivially on structures of cardinality $< \aleph_\alpha$ and as $\mathsf{FO}(Q_\alpha)$ satisfies transfer down to $\aleph_\alpha$, one may concentrate on structures of cardinality $\aleph_\alpha$ when investigating this logic.

For successor ordinals $\alpha + 1$ a reformulation of the problem leads to a new kind of transfer theorems which involve four cardinals. Starting our preparations with the example $\alpha = 0$ we show that the meaning of a sentence of $\mathsf{FO}(Q_1)$ in structures of cardinality $\aleph_1$ can be reduced to the meaning of a related first-order sentence in structures of cardinality $\aleph_1$ where some fixed unary relation symbol U is interpreted by a *countably infinite* set.

Let $\mathcal{A}$ be a τ-structure of cardinality $\aleph_1$. Set $\sigma = \tau \cup \{U, <, F\}$ with new relation symbols U (unary), $<$ (binary), and F (ternary). Expand $\mathcal{A}$ to a σ-structure $\mathcal{B}$ by taking $U^{\mathcal{B}}$ *countably infinite* and letting $(A, <^{\mathcal{B}})$ be an ordering and $F^{\mathcal{B}}$ a ternary relation over A such that for all $a \in A$, $F(\cdot, \cdot, a)$ describes an injection from the initial section of $(A, <^{\mathcal{B}})$ determined by a into $U^{\mathcal{B}}$. Then $(A, <^{\mathcal{B}})$ is an $\aleph_1$-like ordering, and an arbitrary subset of A is uncountable just in case it is unbounded in $(A, <^{\mathcal{B}})$. Hence, for example, $\neg Qy\, Py$ (with unary P in τ) has the same meaning in $\mathcal{A}$ as $\exists z \, \forall y \, (Py \rightarrow y < z)$ has in $\mathcal{B}$.

Now the question whether, for instance, $\mathsf{FO}(Q_{\alpha+1})$ has the same valid sentences as $\mathsf{FO}(Q_1)$ or satisfies the compactness theorem for countable sets of sentences like $\mathsf{FO}(Q_1)$ does, can be reduced to the question whether any countable set of first-order sentences with a designated unary relation symbol U which has a model $\mathcal{A}$ of cardinality $\aleph_{\alpha+1}$ with $|U^{\mathcal{A}}| = \aleph_\alpha$, has a model $\mathcal{B}$ of cardinality $\aleph_1$ with $|U^{\mathcal{B}}| = \aleph_0$, and vice versa. Writing

$$(\aleph_{\alpha+1}, \aleph_\alpha) \rightarrow (\aleph_1, \aleph_0) \text{ and } (\aleph_1, \aleph_0) \rightarrow (\aleph_{\alpha+1}, \aleph_\alpha)$$

for the respective transfer property, we arrive at the question for what cardinals $\aleph_0 \leq \mu_i \leq \kappa_i$ $(i = 1, 2)$ we have $(\kappa_1, \mu_1) \rightarrow (\kappa_2, \mu_2)$.

Respective results are known as *two-cardinal theorems.* We list some examples. Further ones, also examples involving even more cardinals, can be found in [Schmerl, 1985]. For a systematic treatment and a representation of the combinatorial methods involved see [Devlin, 1984; Kanamori, 1994], and also [Chang and Keisler, 1973].

THEOREM.

(a) For all ordinals α, $(\aleph_{\alpha+1}, \aleph_\alpha) \rightarrow (\aleph_1, \aleph_0)$.

(b) (GCH) For all ordinals α with $\aleph_\alpha$ regular, $(\aleph_1, \aleph_0) \rightarrow (\aleph_{\alpha+1}, \aleph_\alpha)$.

(c) $(V = L)$ For all ordinals α, $(\aleph_1, \aleph_0) \rightarrow (\aleph_{\alpha+1}, \aleph_\alpha)$.

(d) $(V = L)$ For all ordinals α and β, $(\aleph_{\alpha+2}, \aleph_\alpha) \rightarrow (\aleph_{\beta+2}, \aleph_\beta)$.

Part (a) is Vaught's Two-Cardinal Theorem from [Morley and Vaught, 1962], part (b) Chang's Two-Cardinal Theorem from [Chang, 1965]. The assumption of the Generalized Continuum Hypothesis in (b) is necessary because the proof uses saturated models. The extension (c) of (b) under the Axiom of Constructibility is due to Jensen [1972]; its proof replaces the use of saturated models by that of special ones. Parts (a) to (c) are examples of gap-1 theorems. Part (d), also due to Jensen, is an example of a gap-2 theorem.

According to, say, part (c) we have under $V = L$ that for all ordinals α, the logic $\mathsf{FO}(Q_{\alpha+1})$ has the same valid sentences as $\mathsf{FO}(Q_1)$, satisfies the compactness theorem for countable sets of sentences and also the completeness theorem for countable sets of premises with the same set of rules as that from [Keisler, 1970] for $\mathsf{FO}(Q_1)$. (Compactness and completeness even hold for sets of sentences of cardinality $\leq \aleph_\alpha$; see [Schmerl, 1985].)

We conclude with some simple counterexamples. Using $\aleph_1$-like orderings, one instantly sees that

$$(\aleph_1, \aleph_0) \not\rightarrow (\aleph_2, \aleph_0),$$

and the construction in the preceding subsection shows that

$$\text{for all } n \in \omega,\ \big(\beth_n(\aleph_0), \aleph_0\big) \not\rightarrow \big(\beth_{n+1}(\aleph_0), \aleph_0\big).$$

However, in Subsection 6. 4 we shall see that

$$\big(\beth_\omega(\aleph_0), \aleph_0\big) \rightarrow \big(\kappa, \aleph_0\big) \text{ for all } \kappa \geq \beth_\omega(\aleph_0).$$

6 TRANSFER IN GENERAL LOGIC SYSTEMS

As emphasized earlier, the Löwenheim-Skolem Theorem witnesses a weakness of first-order logic. On the one hand, first-order model theory lives on this weakness. For example, the characterization of the models of a given theory up to elementary equivalence (in the sense of satisfying the same first-order sentences) may provide valuable insights in cases where a characterization up to isomorphism is still missing. On the other hand, this weakness limits the applicability of first-order model theory as the structures under consideration may not be first-order characterizable. Using stronger logics to overcome this shortage may result in reduced model-theoretic possibilities. The 1970s brought the rise of a general theory for investigating the model-theoretic behaviour of logics which allows to facilitate tracing model-theoretically promising systems [Barwise and Feferman, 1985]. In the present section and the next one we report on some of its results that deal with Löwenheim-Skolem-type phenomena.

6.1 General logic systems

Up to now we have met a variety of logics: first-order logic, (weak) second-order logic, logics with cardinality quantifiers, infinitary logics. The notion of a general logic system is designed to catch essential features of these examples. We provide only a rough description, referring for details to [Ebbinghaus, 1985].

A *general logic system* L is given by

- a mapping defined on vocabularies which maps a vocabulary τ to a class $\mathsf{L}[\tau]$, the class of τ-*sentences* of L,
- a binary relation $\models_{\mathsf{L}}$, the L-*satisfaction relation,* which consists of pairs $(\mathcal{A}, \varphi)$ where – for suitable τ – $\mathcal{A}$ is a τ-structure and $\varphi \in \mathsf{L}[\tau]$,

such that the following properties hold:

(a) $\tau \subseteq \sigma \ \rightarrow\ \mathsf{L}[\tau] \subseteq \mathsf{L}[\sigma]$.

(b) Isomorphic τ-structures satisfy the same sentences of $\mathsf{L}[\tau]$.

(c) For $\tau \subseteq \sigma$, $\varphi \in \mathsf{L}[\tau]$, and $\mathcal{A}$ a σ-structure, $(\ \mathcal{A} \models_{\mathsf{L}} \varphi$ iff $\mathcal{A} \upharpoonright \tau \models_{\mathsf{L}} \varphi)$.

(d) The satisfaction relation $\models_{\mathsf{L}}$ is invariant under renamings of symbols.

A general logic system L is *small* if every $\mathsf{L}[\tau]$ is a set. L has the *finite occurrence property* if each sentence of L belongs to $\mathsf{L}[\tau]$ for some finite τ, and it is of *bounded occurrence* if there is a cardinal ν such that each sentence of L belongs to $\mathsf{L}[\tau]$ for some τ with $|\tau| \leq \nu$. L is *compact* if it obeys the analogue of the compactness theorem for arbitrary sets of sentences, and it is *countably compact* if it does so for countable sets of sentences.

If appropriate, we admit variables in a natural way.

The general logic systems we have mentioned so far, are *regular* in the following sense: They contain all atomic sentences, are closed under the usual propositional connectives, first-order quantifications, relativizations, and the substitution of relation symbols by defined relations of the same arity, and, moreover, allow to eliminate function symbols and constants by use of relation symbols (namely their graphs). Any regular general logic system is at least as expressive as first-order logic.

For simplicity, we shall speak of *logics* when referring to regular general logic systems.

6.2 Löwenheim numbers

The preceding sections present a variety of results about downward and upward transfer for various logics. There is a uniform kind of measuring the extent to which such Löwenheim-Skolem phenomena are valid in a logic, namely by two types of cardinals, Löwenheim numbers as a measure for downward transfer and Hanf numbers as a measure for upward transfer. We start with Löwenheim numbers.

DEFINITION. The *Löwenheim number* $l(\mathsf{L})$ of a logic L is the least cardinal μ such that every satisfiable sentence of L has a model of cardinality $\leq \mu$, provided such a cardinal exists; otherwise, $l(\mathsf{L}) = \infty$.[3]

Using inequalities between constants, one gets $l(\mathsf{L}) \geq \aleph_0$ for any logic L and, in particular, $l(\mathsf{L}_{\kappa^+\omega}) \geq \kappa$ for all infinite cardinals κ. Moreover, $l(\mathsf{FO}(Q_\alpha)) \geq \aleph_\alpha$ for all ordinals α. Downward transfer yields

$$l(\mathsf{FO}) = \aleph_0,\ l(\mathsf{FO}(Q_\alpha)) = \aleph_\alpha,\ \text{and}\ l(\mathsf{L}_{\kappa^+\omega}) = \kappa.$$

Let $\mathsf{L}_{\infty\omega} := \bigcup_{\kappa\geq\omega} \mathsf{L}_{\kappa^+\omega}$ be the union of all $\mathsf{L}_{\kappa^+\omega}$. Obviously, $l(\mathsf{L}_{\infty\omega}) = \infty$ and $\mathsf{L}_{\infty\omega}$ is neither small nor of bounded occurrence. On the other hand, we have the following existence result which may be regarded as a general downward transfer theorem:

THEOREM. Let L be a small logic of bounded occurrence. Then $l(\mathsf{L}) < \infty$.

Proof. Let ν witness the bounded occurrence of L. By the renaming property (d) of general logic systems it suffices to consider a fixed vocabulary τ_0 which consists of ν many relation symbols of any arity n, ν many function symbols of any arity n, and ν many constants. For a satisfiable τ_0-sentence φ of L, let $\kappa(\varphi)$ be the least cardinality of a model of φ. Then

$$l(\mathsf{L}) \ =\ \mathsf{sup}\big\{\kappa(\varphi) \mid \varphi \text{ is a satisfiable sentence of } \mathsf{L}[\tau_0]\big\}.$$

The supremum is $< \infty$ as $\mathsf{L}[\tau_0]$ is a set. ■

The proof is a pure existence proof. In fact, a description of the Löwenheim number may be impossible already for familiar logics as shown by Barwise [1972a] for second-order logic SO.

6.3 Hanf numbers

DEFINITION. The *Hanf number* $h(\mathsf{L})$ of a logic L is the least cardinal μ such that every sentence of L which has a model of cardinality $\geq \mu$, has arbitrarily large models, provided such a cardinal exists; otherwise, $h(\mathsf{L}) = \infty$.[4]

By regularity of L we have $h(\mathsf{L}) \geq \aleph_0$. Moreover, $h(\mathsf{FO}) = \aleph_0$ because of the Upward Löwenheim-Skolem-Tarski Theorem. In Subsection 5. 1 we showed that $h(\mathsf{FO}(Q_1)) \geq \beth_\omega(\aleph_0)$. A similar argument yields that in case $h(\mathsf{L})$ exists, $h(\mathsf{L}) = \aleph_0$

---

[3] Given a cardinal κ, the *κ-Löwenheim number* $l_\kappa(\mathsf{L})$ is defined similarly, referring to sets of sentences of L of cardinality $\leq \kappa$ instead of single sentences. By definition, $l(\mathsf{L}) = l_1(\mathsf{L})$. As a rule, results about $l(\mathsf{L})$ can be suitably generalized to $l_\kappa(\mathsf{L})$. We shall concentrate here on $l(\mathsf{L})$ (and similarly will do so in respect to Hanf numbers and well-ordering numbers introduced below).

[4] It can be shown that $h(\mathsf{L})$ is the supremum of all cardinals μ such that there is a sentence of L which has a model of cardinality μ, but not arbitrarily large models.

or $h(\mathsf{L})$ is a limit cardinal, i. e., of the form $\aleph_\alpha$ for some limit ordinal α. By considering

$$\varphi := \forall x \bigvee \{x = c_\alpha \mid \alpha < \kappa\}$$

we get $h(\mathsf{L}_{\kappa^+\omega}) \geq \kappa^+$ for infinite κ and, hence, $h(\mathsf{L}_{\infty\omega}) = \infty$. On the other hand, an argument like that for the preceding theorem for Löwenheim numbers shows the following observation by Hanf [1960]:

THEOREM. Let L be a small logic of bounded occurrence. Then $h(\mathsf{L}) < \infty$.

As in the case of Löwenheim numbers the theorem leaves us with nothing more than existence. The next subsection will provide some explicit results.

What about the relationship between Löwenheim numbers and Hanf numbers? We have

$$l(\mathsf{FO}) = h(\mathsf{FO})$$

and, as shown below,

$$l(\mathsf{L}) < h(\mathsf{L}) \text{ for } \mathsf{L} = \mathsf{FO}(Q_\alpha),\ \mathsf{L}_{\kappa^+\omega}, \text{ and } \mathsf{WSO}.$$

Furthermore, there is a natural logic L such that the statement

$$h(\mathsf{L}) < l(\mathsf{L})$$

is consistent with ZFC (cf. [Väänänen, 1982]). So there is no general relationship. For more details see the literature mentioned in [Mundici, 1985].

6.4 Hanf numbers and well-ordering numbers

So far we have obtained upward transfer results only for first-order logic, thereby making essential use of the compactness theorem. But incompactness does not necessarily exclude upward transfer. Below we report on some kind of substitute for compactness which as a rule corresponds to reasonably small Hanf numbers. The following fact forms the intuitive starting point.

PROPOSITION. A logic L is not countably compact iff there is a vocabulary τ containing $<$ and a countable set Φ of τ-sentences of L such that the class

$$\{<^{\mathcal{A}} \mid \mathcal{A} \text{ is a } \tau\text{-structure with } \mathcal{A} \models_{\mathsf{L}} \Phi\}$$

is the class of well-ordering relations of type ω.

Proof. For the direction from left to right — the other one is obvious — assume that L is not countably compact and let $\Psi = \{\psi_0, \psi_1, \psi_2, \ldots\}$ be a countable set of τ-sentences of L for which every finite subset has a model, but Ψ itself has no model. As L is regular, τ may be chosen relational. Using new binary relation symbols $<$ and R, let the (countable) set Φ consist of the following sentences:

(a) "$<$ is an ordering relation without a last element";

(b) $\forall x \big(\text{“}x \in \text{ field}(<)\text{”} \rightarrow \exists z\, Rxz\big)$;

(c) $\forall x \big(\text{“}x \in \text{ field}(<)\text{”} \rightarrow \big(\text{“}|\{y \mid y < x\}| \geq n\text{”} \rightarrow \psi_n^{\{z \mid Rxz\}}\big)\big)$ for all n.

We show that Φ satisfies the right side. As each finite subset of Ψ has a model, Φ has a model where $<$ is interpreted by an ordering relation of type ω. As Ψ is unsatisfiable, no model $\mathcal{B}$ of Ψ has an element with infinitely many $<^{\mathcal{B}}$-predecessors. Hence, in any model $\mathcal{B}$ the ordering relation $<^{\mathcal{B}}$ is of order type ω. ■

Intuitively, the proposition exemplifies that compactness corresponds to the noncharacterizability of well-orderings. In contrast to the simple alternative "compact – not compact" the capability of a logic to characterize infinite well-orderings allows a lot of different degrees which correspond to the amount of infinite well-orderings characterizable in it. This amount is made precise by the notion of well-ordering number.

DEFINITION. Let L be a logic.

(a) Given τ with $< \in \tau$, a τ-sentence φ of L *pins down* the ordinal α if for all models $\mathcal{A}$ of φ the relation $<^{\mathcal{A}}$ is a well-ordering relation, and there is a model $\mathcal{A}$ of φ such that $<^{\mathcal{A}}$ is a well-ordering relation of type α.

(b) L is *bounded* if no sentence of L pins down arbitrarily large ordinals.

(c) The *well-ordering number* $w(\mathsf{L})$ of L is the supremum of all ordinals α which are pinned down by sentences of L, provided the supremum exists; otherwise, $w(\mathsf{L}) = \infty$.

Obviously, all ordinals $< w(\mathsf{L})$ can be pinned down by sentences of L. Hence, $w(\mathsf{L})$ is an adequate measure for the amount of well-orderings characterizable in L.

By regularity of L we have $w(\mathsf{L}) \geq \omega$. Therefore, compactness yields that

$$w(\mathsf{FO}) = w(\mathsf{FO}(Q_1)) = \omega.$$

In $\mathsf{L}_{\omega_1\omega}$ all countable ordinals $\alpha \neq 0$ can be pinned down in a strong way, namely by $\{<\}$-sentences φ_α which axiomatize the class of well-orderings of type α. Setting $\varphi_0 := \neg\exists x\, x = x$, they can inductively be defined as follows:

$$\begin{aligned}\varphi_\alpha \quad := \quad & < \text{ is an ordering of the universe} \\ & \wedge \forall x \bigvee \big\{\varphi_\beta^{\{z \mid z < x\}} \mid \beta < \alpha\big\} \wedge \bigwedge \big\{\exists x\, \varphi_\beta^{\{z \mid z < x\}} \mid \beta < \alpha\big\}.\end{aligned}$$

Hence, $w(\mathsf{L}_{\omega_1\omega}) \geq \aleph_1$. Similarly, $w(\mathsf{L}_{\kappa^+\omega}) \geq \kappa$ for all infinite κ and, hence, $w(\mathsf{L}_{\infty\omega}) = \infty$. More exactly it can be shown that $w(\mathsf{L}_{\omega_1\omega}) = \aleph_1$ and that $\mathsf{L}_{\infty\omega}$ is bounded ([Morley, 1965] and [López-Escobar, 1966], respectively). On the other hand, second-order logic is not bounded as it admits a characterization of the class of all well-orderings, i.e., the class of all orderings where each nonempty subset has a least element.

As we shall see now, large well-ordering numbers lead to large Hanf numbers. Let μ be a cardinal and L a logic having a sentence φ which has a model of cardinality μ, but no larger one. An argument similar to that for $\mathsf{FO}(Q_1)$ in Subsection 5. 1 yields

$$h(\mathsf{L}) \geq \beth_\omega(\mu).$$

However, we get more: Let $w(\mathsf{L}) < \infty$. The inductive step in the argument which leads from the sentence φ_n having a model of cardinality $\beth_n(\mu)$ but no larger one to the sentence φ_{n+1} having a model of cardinality $\beth_{n+1}(\mu)$ but no larger one, can be iterated along ordinals which are pinned down by L-sentences. We thus obtain:

THEOREM. Let L be a logic with $w(\mathsf{L}) < \infty$ and φ an L-sentence having a model of cardinality μ, but no larger one. Then

$$h(\mathsf{L}) \;\geq\; \beth_{w(\mathsf{L})}(\mu).$$

In particular, $h(\mathsf{L}_{\omega_1\omega}) \geq \beth_{\aleph_1}(\aleph_0)$ and again $h(\mathsf{FO}(Q_1)) \geq \beth_\omega(\aleph_0)$.

In many cases, the reverse inequality is true, too. As an example we sketch the argument for $\mathsf{FO}(Q_1)$. The result is due to Vaught; cf. [Morley, 1965].

In Subsection 5. 2 we saw that in structures of cardinality $\aleph_1$ sentences of $\mathsf{FO}(Q_1)$ can be replaced by first-order sentences if an $\aleph_1$-like ordering is present. By minor modifications this possibility can be extended to arbitrary uncountable structures. More exactly, we set

$$\mathcal{O} \;:=\; \text{the class of } \aleph_1\text{-like orderings}$$

and define $\mathsf{FO}(\mathcal{O})$, *first-order logic with standard part* $\mathcal{O}$, by admitting only vocabularies τ containing the relation symbols U (unary) and $<$ (binary) and allowing only structures $\mathcal{A}$ where $<^{\mathcal{A}} \subseteq U^{\mathcal{A}} \times U^{\mathcal{A}}$ and $(U^{\mathcal{A}}, <^{\mathcal{A}}) \in \mathcal{O}$. In order to get

$$h(\mathsf{FO}(Q_1)) \;\leq\; \beth_\omega(\aleph_0),$$

it suffices to show

$$h(\mathsf{FO}(\mathcal{O})) \;\leq\; \beth_\omega(\aleph_0).$$

The proof of the latter inequality uses the techniques of indiscernibles described in Subsection 3. 4, but in a more elaborated way as the indiscernibles have to respect the standard part from $\mathcal{O}$ of the structures involved. In order to master these difficulties, one uses the Erdös-Rado Theorem instead of Ramsey's Theorem. The higher cardinalities needed are at hand because one may work with models of cardinality $\beth_\omega(\aleph_0)$.

Altogether we now have

$$h(\mathsf{FO}(Q_1)) \;=\; \beth_\omega(\aleph_0).$$

The method applies to other logics treated so far and yields, for example,

$$h(\mathsf{L}_{\omega_1\omega}) \;=\; \beth_{\aleph_1}(\aleph_0)$$

[López-Escobar, 1966]. We finally mention the following result of Morley [1967]:

$$w(\mathsf{WSO}) = w(\mathsf{FO}(Q_0)) = \omega_1^c \text{ and } h(\mathsf{WSO}) = h(\mathsf{FO}(Q_0)) = \beth_{\omega_1^c}(\aleph_0),$$

where ω_1^c is the first non-recursive ordinal. For more information cf. [Flum, 1972; Flum, 1975; Barwise, 1975; Nadel, 1985; Dickmann, 1985].

Logics L such as second-order logic where $h(\mathsf{L}) < \infty$ and the class of all well-orderings is definable, satisfy $h(\mathsf{L}) = \beth_{h(\mathsf{L})}(0)$ [Barwise, 1974]. However, in general results about the Hanf number of logics which are not bounded are rare. For a structural explanation of this fact cf. [Baldwin, 1985] and the introductory hints in [Barwise, 1985].

7 GENERAL ASPECTS

We conclude the considerations about logics by two examples which illustrate the scope of the Löwenheim-Skolem Theorem in a general context.

7.1 Lindström-type theorems

For logics L_1 and L_2 we write $\mathsf{L}_1 \preceq \mathsf{L}_2$ if for every τ and every $\varphi_1 \in \mathsf{L}_1[\tau]$ there is $\varphi_2 \in \mathsf{L}_2[\tau]$ such that φ_1 and φ_2 have the same models. $\mathsf{L}_1 \equiv \mathsf{L}_2$ means that $\mathsf{L}_1 \preceq \mathsf{L}_2$ and $\mathsf{L}_2 \preceq \mathsf{L}_1$. Intuitively, $\mathsf{L}_1 \equiv \mathsf{L}_2$ if L_1 and L_2 have the same expressive power. For all logics L we have $\mathsf{FO} \preceq L$ by regularity.

In the introduction we addressed already one of Lindström's theorems from [Lindström, 1969]. Essentially it says that first-order logic FO is the strongest logic which admits a complete axiomatization and obeys the Löwenheim-Skolem Theorem. So there is no hope to realize these methodologically valuable properties in a logic with greater expressive power. In the following we give two further theorems from Lindström's paper where completeness is replaced by compactness and by upward transfer, respectively.

THEOREM. Let L be a logic with the finite occurrence property.

(a) If $l(\mathsf{L}) = \aleph_0$ and L is countably compact, then $\mathsf{L} \equiv \mathsf{FO}$.

(b) If $l(\mathsf{L}) = \aleph_0$ and $h(\mathsf{L}) = \aleph_0$, then $\mathsf{L} \equiv \mathsf{FO}$.

H. Friedman [1970] observed that the finite occurrence property is not needed for (a) if L is (fully) compact, as in a compact logic the meaning of a sentence depends only on finitely many of its symbols. (For a proof cf., e. g., [Ebbinghaus *et al.*, 1994]. See [Shelah and Väänänen, 2005] for possible extensions of (a) to infinitary logics.) Part (b) demonstrates in a convincing way the power which is unfolded by the simplest forms of upward and downward transfer.

We give the idea of a *common proof* for parts (a) and (b) in order to demonstrate a methodological relationship between upward transfer and compactness in case the Löwenheim number is $\aleph_0$.

Let L be a logic having the finite occurrence property and Löwenheim number $\aleph_0$. Assume that L is strictly stronger than first-order logic, and let τ be finite and $\varphi \in \mathsf{L}[\tau]$ a sentence which is not equivalent to a sentence in $\mathsf{FO}[\tau]$. Because of regularity τ may be chosen relational.

Claim. There are a finite relational vocabulary $\sigma \supseteq \tau$, a binary relation symbol $E \in \sigma \setminus \tau$, and a satisfiable σ-sentence ψ of L such that for all σ-structures $\mathcal{B}$ with $\mathcal{B} \models \varphi$, the reduct $(B, E^{\mathcal{B}})$ is isomorphic to $(\omega, <)$, i. e., the class of well-orderings of type ω is a so-called pseudoelementary class of L.

By the claim, L is not countably compact and its Hanf number is greater than $\aleph_0$; so the proofs of (a) and (b) are complete.

In order to prove the claim, one first shows that for any $m \geq 1$ there are countable τ-structures $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A} \models_{\mathsf{L}} \varphi$, $\mathcal{B} \models_{\mathsf{L}} \neg\varphi$, and $\mathcal{A}$ and $\mathcal{B}$ satisfy the same first-order τ-sentences of quantifier rank $\leq m$. By a theorem of Fraïssé this yields that for all $m \geq 1$ there are countable τ-structures $\mathcal{A}$ and $\mathcal{B}$ with

$$(*)_m \qquad \mathcal{A} \models_{\mathsf{L}} \varphi,\ \mathcal{B} \models_{\mathsf{L}} \neg\varphi,\ \text{ and } \mathcal{A} \cong_m \mathcal{B}.^5$$

As $\cong_\omega$ yields $\cong$ on countable structures, there are *no* countable τ-structures $\mathcal{A}$ and $\mathcal{B}$ with $(*)_\omega$.

Regarding partial isomorphisms as first-order objects, one can describe this situation in L and in this way obtain $\sigma \supseteq \tau$, $E \in \sigma \setminus \tau$, and a satisfiable σ-sentence ψ of L such that ψ ensures for any model $\mathcal{B}$ that $(B, E^{\mathcal{B}})$ is a discrete ordering without a last element which allows descending sequences of length $\alpha \in \omega \cup \{\omega\}$ just in case $\alpha \in \omega$, i. e., that $(B, <^{\mathcal{B}})$ is of type ω.

7.2 First-order-like logics

There is a principal difference between first-order logic and second-order logic: In FO the satisfaction relation is "of local character" in the sense that the truth or falsity of $\mathcal{A} \models \varphi$ is essentially fixed by φ and its subformulas and by $\mathcal{A}$ and its elements. In SO, however, it may depend on the whole set-theoretic universe, in particular, on what subsets of A there are. For example, it is easy to give a second-order sentence which is valid just in case Cantor's Continuum Hypothesis holds. The local character of the first-order satisfaction relation may be made more precise by demanding that it should not depend on changes of the set-theoretic universe "far apart", i. e., it should be invariant under *end extensions* of the universe, extensions where old sets do not get new elements.

We shall provide an impression of these ideas, ending with (a special case of) a general downward Löwenheim-Skolem type theorem. The theory goes back to Barwise [1972b] with an extended framework given in [Barwise, 1975].

---

[5] Given structures $\mathcal{A}$ and $\mathcal{B}$ of the same relational vocabulary, a *partial isomorphism* from $\mathcal{A}$ into $\mathcal{B}$ is the empty mapping or an isomorphism from a substructure of $\mathcal{A}$ onto a substructure of $\mathcal{B}$. For $m \in \omega$, $\mathcal{A} \cong_m \mathcal{B}$ means that the empty mapping can be extended m times, in each step prescribing either an element from A for the domain or an element from B for the range, such that the resulting mapping is a partial isomorphism from $\mathcal{A}$ into $\mathcal{B}$. $\mathcal{A} \cong_\omega \mathcal{B}$ is defined analogously.

We start with some definitions.

A first-order formula $\varphi(\overline{x})$ of set theory is

- *bounded* if all its quantifiers are restricted to sets;
- Σ_1 if it has the form $\exists \overline{y}\, \psi$ where ψ is bounded;
- Σ_1^{ZFC} if it is equivalent in ZFC to a Σ_1 formula.

The formula $\varphi(\overline{x})$ is *persistent* if for all models $\mathcal{A}$ and $\mathcal{B}$ of ZFC where $\mathcal{B}$ is an end extension of $\mathcal{A}$, we have for all $\overline{a} \in A$ that

$$\mathcal{A} \models \varphi(\overline{a}) \text{ implies } \mathcal{B} \models \varphi(\overline{a}).$$

By a theorem of Feferman and Kreisel [1966], $\varphi(\overline{x})$ is persistent iff it is Σ_1^{ZFC}.

Vocabularies and structures can be defined set-theoretically. Moreover, this can be done in such a manner that the predicate "x is a vocabulary and y is an x-structure" is Σ_1^{ZFC}. All the logics treated so far can be given in set-theoretic terms, too. According to the introductory remarks above it seems appropriate to make precise the notion of being a first-order like logic by imposing conditions of persistence on (the set-theoretic definition of) the syntax and the semantics.

Following this idea, we call a logic L *weakly absolute* if the set-theoretic predicate "$x \in \mathsf{L}[\tau]$" and the satisfaction relation are persistent and, hence, Σ_1^{ZFC}.[6]

FO, WSO, FO(Q_0), and $\mathsf{L}_{\omega_1\omega}$ are weakly absolute, whereas the following theorem will show that FO(Q_1) and SO are not.

Before giving it, we still need a crucial fact from ZFC set theory. The *transitive closure* $\mathsf{TC}(x)$ of a set x is the smallest transitive set $y \supseteq x$, i. e., the smallest set $y \supseteq x$ such that each element of y is a subset of y. Let

$$||x|| := \mathsf{max}\{\aleph_0, |\mathsf{TC}(x)|\}.$$

Then an argument based on the Löwenheim-Skolem Theorem for first-order sentences of set theory yields the following result of Levy [1965]:

LEMMA. If $\varphi(x, y, z)$ is Σ_1^{ZFC}, then ZFC proves

$$\exists x\, \varphi(x, y, z) \;\rightarrow\; \exists x\big(||x|| \leq \mathsf{max}\{||y||, ||z||\} \wedge \varphi(x, y, z)\big).$$

THEOREM. Let L be a weakly absolute logic and $\varphi \in \mathsf{L}[\tau]$. If φ is satisfiable, it has a model of cardinality $\leq \mathsf{max}\{||\varphi||, ||\tau||\}$.

Proof. As the conjunction of two Σ_1^{ZFC} formulas is Σ_1^{ZFC}, a weakly absolute logic L admits a Σ_1^{ZFC} formula $S(x, y, z)$ such that for all vocabularies τ, all sentences $\varphi \in \mathsf{L}[\tau]$, and all sets x,

$$S(x, \varphi, \tau) \text{ iff } x \text{ is a } \tau\text{-structure and } x \models_{\mathsf{L}} \varphi.$$

---

[6] In general, the negation of a persistent formula is not persistent, i. e., persistent formulas need not be (fully) invariant or *absolute* under end extensions. *Fully absolute* logics impose conditions of absoluteness on the satisfaction relation.

By assumption there is a τ-structure $\mathcal{A}$ with $\mathcal{A} \models_{\mathsf{L}} \varphi$. We, hence, have

$$\exists\, \mathcal{A}\, S(\mathcal{A}, \varphi, \tau).$$

As the set-theoretic definition of structures ensures $A \subseteq TC(\mathcal{A})$, the lemma yields the claim. ■

For the weakly absolute logics FO, WSO, $\mathsf{FO}(Q_0)$, and $\mathsf{L}_{\omega_1\omega}$ the vocabularies τ may be chosen countable, and the usual set-theoretic definitions are given in such a way that $||\varphi||, ||\tau|| \leq \aleph_0$. We thus get back all our results about transfer down to $\aleph_0$. In this sense the theorem may be regarded as a general form of a first-order like Löwenheim-Skolem Theorem for sentences. For its general version and a comprehensive presentation of the theory of absolute logics, see also [Väänänen, 1985].

8 TRANSFER IN THE FINITE

Most results and techniques of first-order model theory fail or get trivial if one considers only finite structures. Interpolation and preservation theorems and the method of ultraproducts are prominent examples. Nevertheless, some parts such as the Ehrenfeucht-Fraïssé method survive (cf. [Ebbinghaus and Flum, 1995; Libkin, 2004]). In the present section we consider Löwenheim-Skolem phenomena. It should come with no surprise that positive results are rare, whereas negative ones are abundant.

Given a finite vocabulary τ, the following trivial facts correspond to a weak form of the existence of Löwenheim numbers and Hanf numbers:

(Lfin) There exists a function $l : \mathsf{FO}[\tau] \to \omega$ such that every τ-sentence φ which has a finite model, has a model of cardinality $\leq l(\varphi)$.

(Hfin) There exists a function $h : \mathsf{FO}[\tau] \to \omega$ such that every τ-sentence φ which has a finite model of cardinality $\geq h(\varphi)$, has arbitrarily large finite models.

They turn into interesting questions if one imposes conditions on l and h, say, by asking whether these functions can be chosen recursive, allowing that they are defined only on certain subsets of $\mathsf{FO}[\tau]$, and/or by letting them depend only on certain parameters such as length, quantifier rank, or number of variables.

Positive results can be extracted from proofs of the finite model property. We mention the *Bernays-Schönfinkel classes*

$$\exists^*\forall^*\,[\tau] := \{\varphi \in \mathsf{FO}[\tau] \mid \varphi = \exists x_1 \ldots \exists x_{n_\varphi} \forall y_1 \ldots \forall y_{m_\varphi} \psi_\varphi,\ \psi_\varphi \text{ quantifier-free}\}$$

for relational τ. Using the fact that model classes of universal sentences are closed under substructures, one gets immediately that on $\exists^*\forall^*\,[\tau]$ the function

$$l(\varphi) := n_\varphi$$

can serve for (Lfin). Ramsey's proof [Ramsey, 1930] for the decidability of the satisfiability problem for $\exists^*\forall^*\,[\tau]$ yields that (Hfin) can be witnessed by a function of the form

$$h(\varphi) \;=\; h_2(n_\varphi, m_\varphi),$$

where h_2 is primitive recursive on $\omega \times \omega$.

There is no hope for a substantial extension of these results. For the arithmetical vocabulary $\sigma := \{0, +, \times, <, S\}$ (with S for the successor relation) and for "reasonable" Gödel numberings $\varphi \longmapsto \ulcorner\varphi\urcorner$ on $\exists^*\forall^*\,[\sigma]$, Grohe [2002] showed:

THEOREM. (a) For each recursive function l there exist $\varphi \in \exists^*\forall^*\,[\sigma]$ and $n > l(\ulcorner\varphi\urcorner)$ such that φ has a model in all cardinalities $\geq n$, but no model of cardinality $< n$.

(b) For each recursive function h there exist $\varphi \in \exists^*\forall^*\,[\sigma]$ and $n > h(\ulcorner\varphi\urcorner)$ such that φ has a model in all cardinalities $< n$, but no model of cardinality $\geq n$.

Dawar [1998] treats a different version of Löwenheim-Skolem-type theorems in the finite. They refer to the finite variable languages FO^k ($k = 1, 2, 3, \ldots$) where only those first-order sentences are allowed which use at most k distinct variables. In many cases full first-order logic is too strong for finite model theory as each structure (of finite vocabulary) is characterizable up to isomorphism by a single sentence, whereas experience shows that finite variable logics are methodologically well-adapted.

Given a finite relational vocabulary τ, some $k \geq 1$, and a τ-structure $\mathcal{A}$, a *k-type* of $\mathcal{A}$ is a maximal set of formulas $\varphi(x_1, \ldots x_k)$ of $\mathsf{FO}^k[\tau]$ which is satisfiable in $\mathcal{A}$. It can be completely described by one suitably chosen formula. The *k-size* of $\mathcal{A}$ is the number of k-types of $\mathcal{A}$. The k-size reflects the richness of $\mathcal{A}$ as far as it can be expressed in $\mathsf{FO}^k[\tau]$. Instead of considering models of a given τ-sentence of FO^k, Dawar considers τ-structures satisfying the same sentences of FO^k, i. e., models of the same complete FO^k-theory (which can also be finitely axiomatized).

Dawar proposes several techniques with indiscernibles in order to attack the following questions:

Lfin-Question. Given τ, is there a recursive function l_k, for every $k \geq 1$, such that, for every finite τ-structure $\mathcal{A}$ with k-size $\leq n$, there is a τ-structure $\mathcal{B}$ such that $|B| \leq l_k(n)$ and $\mathcal{A}$ and $\mathcal{B}$ satisfy the same τ-sentences of FO^k?

Hfin-Question. Given τ, is there a recursive function h_k, for every $k \geq 1$, such that, if $\mathcal{B}$ is a finite τ-structure with k-size $\leq n$ and $|B| \geq h_k(n)$, then there are arbitrarily large finite τ-structures $\mathcal{A}$ such that $\mathcal{A}$ and $\mathcal{B}$ satisfy the same τ-sentences of FO^k?

He provides a solution for the class of finite trees. However, in general the answer to both questions is negative ([Grohe, 2002] for the Lfin-Question, [Barker, 2003] for the Hfin-Question).

In Subsection 3. 3 we saw that the first-order transfer theorems give information about the spectra of first-order theories. The failure of the search for suitable analogues of Löwenheim-Skolem phenomena in the finite leaves us with the problem of providing general features of the spectra $\mathsf{spec}_{\mathsf{fin}}(\varphi)$ of first-order sentences φ in the finite, originally formulated by Scholz [1952]. According to Jones and Selman [1974] the first-order spectra in the finite coincide with the NP-sets of natural numbers, if the numbers are given in unary notation. So they are linked to fundamental questions in complexity theory. For further information see [Börger *et al.*, 1997], for historical remarks about the problem see [Börger, 1984].

ACKNOWLEDGEMENTS

The author appreciates valuable comments from Jörg Flum and Martin Otto.

BIBLIOGRAPHY

[Badesa, 2004] C. Badesa. *The Birth of Model Theory: Löwenheim's Theorem in the Frame of the Theory of Relatives.* Princeton University Press: Princeton, New Jersey, 2004.

[Baldwin, 1985] J. T. Baldwin. Definable second-order quantifiers. In [Barwise and Feferman, 1985, 445–477].

[Barker, 2003] R. Barker. L_k*-Equivalence and Hanf Functions for Finite Structures.* Ph. D. Thesis, University of Oxford, 2003.

[Barwise, 1972a] J. Barwise. The Hanf number of second order logic. *The Journal of Symbolic Logic* **37**, 588–594.1972.

[Barwise, 1972b] J. Barwise. Absolute logics and $L_{\infty\omega}$. *Annals of Mathematical Logic* **4**, 309–340, 1972.

[Barwise, 1974] J. Barwise. Axioms for abstract model theory. *Annals of Mathematical Logic* **7**, 221–265, 1974.

[Barwise, 1975] J. Barwise. *Admissible Sets and Structures. An Approach to Definability Theory.* Springer-Verlag: Berlin, Heidelberg, New York, 1975.

[Barwise, 1985] J. Barwise. Model-theoretic logics: Background and aims. In [Barwise and Feferman, 1985, 3–23].

[Barwise and Feferman, 1985] J. Barwise and S. Feferman, eds. *Model-Theoretic Logics.* Springer-Verlag: New York, Berlin, *et al.,* 1985.

[Börger, 1984] E. Börger. Spektralproblem and completeness of logical decision problems. In: E. Börger, G. Hasenjaeger, and D. Rödding (eds.): *Logic and Machines: Decision Problems and Complexity.* Lecture Notes in Computer Science 171, Springer-Verlag: Berlin, Heidelberg, New York, *et al.*, 333–356, 1984.

[Börger *et al.*, 1997] E. Börger, E. Grädel and Y. Gurevich. *The Classical Decision Problem.* Springer-Verlag: Berlin, Heidelberg, New York, *et al.*, 1997.

[Chang, 1965] C. C. Chang. A note on the two cardinal problem. *Proceedings of the American Mathematical Society* **16**, 1148–1155, 1965.

[Chang and Keisler, 1973] C. C. Chang and H. J. Keisler. *Model Theory.* North-Holland Publishing Company: Amsterdam *et al.,* 1973.

[Dawar, 1998] A. Dawar. Finite models and finitely many variables. In J. A. Makowsky and E. V. Ravve (eds.): *Logic Colloquium '95.* Lecture Notes in Logic 11, Springer-Verlag: Berlin, Heidelberg, New York, *et al.*, 51–65, 1998.

[Devlin, 1984] K. J. Devlin. *Aspects of Constructibiliy.* Springer-Verlag: Berlin, Heidelberg, New York, *et al.,* 1984.

[Dickmann, 1985] M. A. Dickmann. The infinitary languages $\mathcal{L}_{\kappa\lambda}$ and $\mathcal{L}_{\infty\lambda}$. In [Barwise and Feferman, 1985, 318–363].

[Ebbinghaus, 1985] H.-D. Ebbinghaus. Extended logics: The general framework. In [Barwise and Feferman, 1985, 25–76].

[Ebbinghaus and Flum, 1995] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer-Verlag: Berlin, Heidelberg, New York, *et al.,* 1995.

[Ebbinghaus *et al.*, 1994] H.-D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic.* Springer-Verlag: New York, Berlin, *et al.,* 1994.

[Ehrenfeucht and Mostowski, 1956] A. Ehrenfeucht and A. Mostowski. Models of axiomatic theories admitting automorphisms. *Fundamenta Mathematicae* **43**, 50–68, 1956.

[Etessami *et al.*, 2002] K. Etessami, M. Y. Vardi and T. Wilke. First-order logic with two variables and unary temporal logic. *Information and Computation* **179**, 279–295, 2002.

[Feferman and Kreisel, 1966] S. Feferman and G. Kreisel. Persistent and invariant formulas relative to theories of higher order. *Bulletin of the American Mathematical Society* **72**, 480–485, 1966.

[Flum, 1972] J. Flum. Hanf numbers and well-ordering numbers. *Archiv für Mathematische Logik und Grundlagenforschung* **15**, 164–178, 1972.

[Flum, 1975] J. Flum. First-order logic and its extensions. In: G. H. Müller, A. Oberschelp, and K. Potthoff (eds.): *Logic Conference Kiel 1974.* Lecture Notes in Mathematics 499, Springer-Verlag: Berlin, Heidelberg, New York, 248–310, 1975.

[Flum, 1985] J. Flum. Characterizing logics. In [Barwise and Feferman, 1985, 77–120].

[Friedman, 1970] H. M. Friedman. Why first-order logic? Manuscript, Stanford University and the University of Wisconsin, 1970, 16pp.

[Fuhrken, 1964] G. Fuhrken. Skolem-type normal forms for first-order languages with a generalized quantifier. *Fundamenta Mathematicae* **54**, 291–302, 1964.

[Fuhrken, 1965] G. Fuhrken. Languages with added quantifier 'there exist $\aleph_\alpha$ many'. In: J. W. Addison, L. H. Henkin, and A. Tarski (eds.): *The Theory of Models.* North-Holland Publishing Company: Amsterdam, 121–131. 1965.

[Gabbay, 1971] D. M. Gabbay. Expressive functional completeness in tense logic. In: U. Mönnich (ed.): *Aspects of Philosophical Logic.* Reidel: Dordrecht, 91–117, 1971.

[Grädel *et al.*, 1997] E. Grädel, P. Kolaitis and M. Y. Vardi. On the decision problem for two-variable first-order logic. *The Bulletin of Symbolic Logic* **3**, 53–69, 1997.

[Grohe, 1996] M. Grohe. Some remarks on the finite Löwenheim-Skolem theorems. *Mathematical Logic Quarterly* **42**, 569–571, 1996.

[Grohe, 2002] M. Grohe. Large finite models with small L^{k}-types. *Information and Computation* **179**, 250–278,2002.

[Hanf, 1960] W. P. Hanf. Models of languages with infinitely long expressions. In: *Abstracts of Contributed Papers from the First Logic, Methodology and Philosophy of Science Congress, Stanford University 1960.* Vol. 1, 1960, 24.

[Hodges, 1993] W. Hodges. *Model Theory.* Cambridge University Press: Cambridge *et al.,* 1993.

[Jensen, 1972] R. B. Jensen. The fine structure of the constructible hierarchy. *Annals of Mathematical Logic* **4**, 229–308 and 443, 1972.

[Jones and Selman, 1974] N. D. Jones and A. L. Selman. Turing machines and the spectra of first-order formulas. *The Journal of Symbolic Logic* **39**, 139–150, 1974.

[Kanamori, 1994] A. Kanamori. *The Higher Infinite.* Springer-Verlag: Berlin, Heidelberg, New York, *et al.,* 1994.

[Kaufmann, 1985] M. J. Kaufmann. The quantifier 'There exist uncountably many' and some of its relatives. In [Barwise and Feferman, 1985, 123–176].

[Keisler, 1970] H. J. Keisler. Logic with the quantifier 'there are uncountably many'. *Annals of Mathematical Logic* **1**, 1–93, 1970.

[Levy, 1960] A. Levy. Axiom schemata of strong infinity in axiomatic set theory. *Pacific Journal of Mathematics* **10**, 223–238. 1960.

[Levy, 1965] A. Levy. A hierarchy of formulas in set theory. *Memoirs of the American Mathematical Society* **57**, 76 pp, 1965.

[Libkin, 2004] L. Libkin. *Elements of Finite Model Theory.* Springer-Verlag: Berlin, Heidelberg, New York, *et al.*, 2004.

[Lindström, 1969] P. Lindström. On extensions of elementary logic. *Theoria* **35**, 1–11, 1969.

[Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen* **76**, 447–470, 1915.

[López-Escobar, 1966] E. G. K. López-Escobar. On defining well-orderings. *Fundamenta Mathematicae* **59**, 13–21 and 299–300, 1966.

[Mekler, 1977] A. Mekler. Theories with models of prescribed cardinalities. *The Journal of Symbolic Logic* **42**, 251–253, 1977.

[Montague, 1961] R. M. Montague. Fraenkel's addition to the axioms of Zermelo. In: Y. Bar-Hillel, E. I. J. Poznanski, M. O. Rabin, and A. Robinson (eds.): *Essays in the Foundations of Mathematics.* Magnus Press: Jerusalem, 91–114, 1961.

[Morley, 1965] M. D. Morley. Omitting classes of elements. In: J. W. Addison, L. Henkin, and A. Tarski (eds.): *The Theory of Models.* North-Holland Publishing Company: Amsterdam, 265–273, 1965.

[Morley, 1967] M. D. Morley. The Hanf number for ω-logic (abstract). *The Journal of Symbolic Logic* **32**, 437, 1967.

[Morley and Vaught, 1962] M. D. Morley and R. L. Vaught. Homogeneous and universal models. *Mathematica Scandinavica* **11**, 37–57, 1962.

[Mortimer, 1975] M. Mortimer. On languages with two variables. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **21**, 135–140, 1975.

[Mostowski, 1957] A. Mostowski. On a generalization of quantifiers. *Fundamenta Mathematicae* **44**, 12–36, 1957.

[Mundici, 1985] D. Mundici. Other quantifiers: An overview. In [Barwise and Feferman, 1985, 211–233].

[Nadel, 1985] M. Nadel. $\mathcal{L}_{\omega_1\omega}$ and admissible fragments. In [Barwise and Feferman, 1985, 271–316].

[Ramsey, 1930] F. P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society* **30**, 264–286, 1930.

[Schmerl, 1985] J. H. Schmerl. Transfer theorems and their application to logics. In [Barwise and Feferman, 1985, 177–209].

[Scholz, 1952] H. Scholz. Ein ungelöstes Problem in der symbolischen Logik. *The Journal of Symbolic Logic* **17**, 160, 1952.

[Shelah and Väänänen, 2005] S. Shelah and J. Väänänen. A note on extensions of infinitary logic. *Archive for Mathematical Logic* **44**, 63–69, 2005.

[Skolem, 1920] T. Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit und Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Skrifter, Vitenskapsakademiet i Kristiania, Matematisk-Naturvidenskapelig Klasse* **4**, 1–36, 1920. Reprinted in [Skolem, 1970, 103–16]. English translation of the first section in [van Heijenoort, 1967, 252–263].

[Skolem, 1923] T. Skolem. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Wissenschaftliche Vorträge gehalten auf dem fünften Kongress der skandinavischen Mathematiker in Helsingfors vom 4. bis 7. Juli 1922.* Akademiska Bokhandeln, Helsinki, 217–232, 1923. Reprinted in [Skolem, 1970, 137–152]. English translation in [van Heijenoort, 1967, 290–301].

[Skolem, 1929] T. Skolem. Über einige Grundlagenfragen der Mathematik. *Skrifter, Vitenskapsakademiet i Oslo, Matematisk-Naturvidenskapelig Klasse* **4**, 1–49, 1929. Reprinted in [Skolem, 1970, 227–273].

[Skolem, 1970] T. Skolem. *Selected Works in Logic*, ed. by J. E. Fenstad, Universitetsforlaget: Oslo, Bergen, Tromsö, 1970.

[Tarski and Vaught, 1957] A. Tarski and R. L. Vaught. Arithmetical extensions of relational systems. *Compositio Mathematica* **13**, 81–102, 1957.

[Väänänen, 1982] J. Väänänen. Abstract logic and set theory. II. Large cardinals. *The Journal of Symbolic Logic* **47**, 335–346, 1982.

[Väänänen, 1985] J. Väänänen. Set-theoretic definability of logics. In: [Barwise and Feferman, 1985, 599–643].

[van Heijenoort, 1967] J. van Heijenoort. *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press: Cambridge, Mass., 1967.

[Vaught, 1964] R. L. Vaught. The completeness of logic with the added quantifier 'there are uncountably many'. *Fundamenta Mathematicae* **54**, 303–304, 1964.

[Zermelo, 1908] E. Zermelo. Untersuchungen über die Grundlagen der Mengenlehre. I. *Mathematische Annalen* **65** , 261–281, 1908. English translation in [van Heijenoort, 1967, 199–215].

THE MATHEMATICS OF SKOLEM'S PARADOX

Timothy Bays

In 1922, Thoralf Skolem published a paper entitled "Some Remarks on Axiomatized Set Theory." The paper presents a new proof of a model-theoretic result originally due to Leopold Löwenheim and then discusses some philosophical implications of this result. In the course of this latter discussion, the paper introduces a model-theoretic puzzle that has come to be known as "Skolem's Paradox."

Over the years, Skolem's Paradox has generated a fairly steady stream of *philosophical* discussion; nonetheless, the overwhelming consensus among philosophers and logicians is that the paradox doesn't constitute a *mathematical* problem (i.e., it doesn't constitute a real contradiction). Further, there's general agreement as to *why* the paradox doesn't constitute a mathematical problem. By looking at the way first-order structures interpret quantifiers — and, in particular, by looking at how this interpretation changes as we move from structure to structure — we can give a technically adequate "solution" to Skolem's Paradox. So, whatever the philosophical upshot of Skolem's Paradox may be, the mathematical side of Skolem's Paradox seems to be relatively straightforward.

In this paper, I challenge this common wisdom concerning Skolem's Paradox. While I don't argue that Skolem's Paradox constitutes a genuine mathematical problem (it doesn't), I do argue that standard "solutions" to the paradox are technically inadequate. Even on the mathematical side, Skolem's Paradox is more complicated — and quite a bit more interesting — than it's usually taken to be. Further, because philosophical discussions of Skolem's Paradox typically start with an analysis of the paradox's mathematics — and only then examine how the interpretation of this mathematics reveals the paradox's philosophical significance — it is important to get the mathematics itself right before we start in on our philosophy.

From a structural standpoint, this paper breaks into six sections. In section 1, I formulate a simple version of Skolem's Paradox and try to disentangle the roles that set theory, model theory and philosophy play in making it look plausible. In section 2, I sketch a generic solution to Skolem's Paradox — a solution which explains, in rough outline, why no version of the paradox generates a genuine contradiction. §§3–5 examine different ways of "filling out" this generic solution. §3 focuses on the role quantification sometimes plays in Skolem's Paradox and includes a discussion of the so-called "transitive submodel" version of the paradox.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

§§4 and 5 look at some cases where quantification *doesn't* help to explain Skolem's Paradox. Finally, section 6 presents some concluding philosophical reflections.[1]

1 SKOLEM'S PARADOX

In its simplest form, Skolem's Paradox involves a (seeming) conflict between two theorems of modern logic: Cantor's theorem from set theory and the Löwenheim-Skolem theorem from model theory. Cantor's theorem says that there are uncountable sets — sets which are *too big* to be put into one-to-one correspondence with the natural numbers. The Löwenheim-Skolem theorem says that if a countable collection of first-order sentences has a model, then it has a model whose domain is only countable. Skolem's Paradox arises when we note that the standard axioms of set theory are themselves a countable collection of first-order sentences. How can the very axioms which prove the existence of uncountable sets be satisfied by a merely countable model?

This puzzle can be made somewhat more concrete by considering a specific case. Let T be a standard, first-order axiomatization of set theory — say, ZFC. On the assumption that T has a model, the Löwenheim-Skolem theorem ensures that it has a countable model. Call this model $\mathbb{M}$.[2] Now, as $\mathrm{T} \vdash \exists x$ "*x is uncountable*," there must be some $\hat{m} \in \mathbb{M}$ such that $\mathbb{M} \models$ "*$\hat{m}$ is uncountable*." But, as $\mathbb{M}$ itself is only countable, there are only countably many $m \in \mathbb{M}$ such that $\mathbb{M} \models m \in \hat{m}$. On the surface, then, we seem to have a conflict: from one perspective, $\hat{m}$ looks uncountable, while from another perspective, $\hat{m}$ is clearly countable.

---

[1] Let me emphasize that — with the exception of section 6 and some brief philosophical digressions — this paper focuses fairly tightly on the *mathematical* side of Skolem's Paradox. In particular, I don't attempt to survey all the things philosophers have said about the paradox or to assess the various ways the paradox has been used (and abused) in the philosophical literature. For discussion of such topics, I invite the reader to examine [Bays, 2000] (esp. chapter 3), [Hart, 1970], [Myhill, 1967], and [Resnik, 1966]. I also recommend the illuminating exchange between [Benacerraf, 1985] and [Wright, 1985]. For a quite different view, see [Klenk, 1976] or [Putnum, 1980].

That being said, this paper does serve two philosophical purposes. First, many versions of Skolem's Paradox depend on misleading (and/or outright mistaken) presentations of the underlying mathematics. I think, therefore, that a clear exposition of this mathematics — highlighting all the little twists and turns — already does a lot towards "solving" the paradox. Second, I think philosophers have tended to overemphasize the role *quantification* plays in Skolem's Paradox, and that this overemphasis colors most standard assessments of the paradox's philosophical significance. In sections 4–5, I argue that quantification is less important for Skolem's Paradox than many commentators have supposed, and, in section 6, I say a little about the philosophical upshot of de-emphasizing quantification.

A final comment is in order. Throughout the paper, I relegate a lot of technical machinery — particularly concerning the construction of specific models — to the footnotes. Most of this machinery can be skipped without losing the main thread of argument. The reader who is willing to accept technical claims on faith should feel free to bypass this material. All readers should be warned that some footnotes, especially those in sections 4–5, presuppose substantial mathematical background.

[2] Throughout this paper, I use blackboard bold letters to denote models and the corresponding unbolded letters to denote the domains of those models: so, $\mathbb{M}$ is a model and M is its domain, $\mathbb{N}$ is a model and N is its domain, etc. That being said, I will often abuse notation and write things like "$\mathbb{M}$ is countable" or "$m \in \mathbb{M}$" when I really mean that "M is countable" or that $m \in M$; in context, this should never cause any confusion. Finally, unless otherwise specified, all models should be assumed to be for the language of set theory — i.e., the language with "$\in$" as its sole nonlogical primitive.

In exploring this seeming conflict, I want to begin with three preliminary points. First there's at least one sense in which this appearance of conflict is *clearly* misleading. Strictly speaking, $\mathbb{M}$ doesn't understand ordinary English phrases like "x is uncountable," so the sentence "$\mathbb{M} \models$ '*$\hat{m}$ is uncountable*'" makes no literal sense. Literally, what's going on is the following. There is a specific formula in the language of first-order set theory which mathematicians sometimes find it convenient to *abbreviate* by "x is uncountable." If we avoid this abbreviation — and use, say, "$\Omega(x)$" to denote the relevant formula — then the initial appearance of paradox vanishes. The argument of the last paragraph simply shows that there is some $\hat{m} \in \mathbb{M}$ such that:

1. $\mathbb{M} \models \Omega[\hat{m}]$

and, 2. $\hat{m}$ is countable.

Even on the surface, these claims look philosophically innocuous. After all, lots of models satisfy lots of formulas with respect to lots of parameters, and there's no *general* reason to think that these instances of satisfaction have anything to do with countability and uncountability.

Unfortunately, Skolem's Paradox is a bit harder than this. There's a reason mathematicians often abbreviate $\Omega(x)$ by "x is uncountable," and this reason goes a long way toward explaining why 1 and 2 might — even *under* the surface — continue to look paradoxical. Consider the ordinary English sentence "x is uncountable." If asked what this sentence means, a set theorist will say something about the lack of a bijection between x and the natural numbers. If asked about the phrase "is a bijection," she might go on to talk about collections of ordered pairs satisfying certain nice properties. Finally, if asked about the term "ordered pair," she may say something about the ways one can identify ordered pairs with particular sets.

Suppose our set theorist takes this explanatory process to its logical conclusion. By continuing to fill in the details of "x is uncountable," she will eventually obtain a single sentence which uses no phrases other than "equals," "is a member of," "not," "if. . . then," and "there is a set y, such that." Because this sentence is quite long,[3] she may chose to shorten it by abbreviating the above phrases with the symbols $=, \in, \neg, \rightarrow,$ and $\exists y$. Having done so, she will obtain an explication of the ordinary English sentence "x is uncountable" which uses no symbols other than $=, \in, \neg, \rightarrow,$ and $\exists y$ (and, perhaps, some punctuation).

---

[3] It's important to emphasize just how long this sentence really is. Written explicitly, even a simple phrase like "x is a singleton" turns into the following:

> There is a set a such that it is not the case that if a is a member of x, then there exists a set b such that it is not the case that if b is a member of x then b is equal to a.

If we examine marginally more complicated phrases — say, "x is an ordered pair" or "f is a function" — then we get sentences as long as good size paragraphs. Finally, a full explication of the phrase "x is uncountable" will require several (largely incomprehensible) pages to write down explicitly!

At this point, we should notice something interesting: the sentence our set theorist has just produced *looks exactly like* the first-order formula that we've been calling $\Omega(x)$.[4] That is, if we simply compare the syntax of these two expressions on a symbol-by-symbol basis — ignoring any semantic information we may happen to have about them — we will find that they contain *exactly the same symbols* in *exactly the same order.* This explains why set theorists find it so convenient to abbreviate the formula $\Omega(x)$ with the expression "x is uncountable." It also explains why we might continue to find claims 1 and 2 somewhat puzzling: after all, the formula that $\mathbb{M}$ satisfies in 1 looks just like the negation of claim 2 (after, of course, claim 2 has been fully explicated).

This brings me to my second preliminary point. In explicating claim 2, we need to start with an initial interpretive question. When we say that the element $\hat{m} \in \mathbb{M}$ "is countable," do we mean that

I. $\{x \mid x \in \hat{m}\}$ is countable

or do we mean that

II. $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ is countable?

These two interpretations lead to rather different understandings of what's going on in Skolem's Paradox. In particular, although they each require us to put some constraints on our choice of $\mathbb{M}$, they don't require us to put the *same* constraints on this choice. Hence, it's important to get clear about these interpretations — and their associated constraints — before we go any further.[5]

Let me begin with two comments concerning the *difference* between I and II. First, the two interpretations differ *only* in the way they interpret the notion of "membership" vis-a-vis the element $\hat{m}$. Interpretation I assumes that we are interested in the *real* membership relation on $\hat{m}$, while interpretation II assumes that we are interested in whatever relation $\mathbb{M}$ *thinks* is the membership relation on $\hat{m}$ — i.e., in whatever relation on $M \times M$ serves as the interpretation of "$\in$" under the interpretation function of $\mathbb{M}$.

---

[4] A caveat is in order here. There are many different ways of explicating the notion "x is uncountable," depending on how we decide to "code up" basic set theoretic notions — e.g., *ordered pair* or *natural number.* For convenience, I'm assuming our set theorist has made the *same* decisions we made when we formulated $\Omega(x)$. For any particular $\Omega(x)$, there is *an* explication of "x is uncountable" which has the same syntactic form as $\Omega(x)$; so, we don't lose any generality in assuming that it's the explication our set theorist actually came up with (if it isn't, then we can just find another, more accommodating, set theorist!).

Although this issue about coding is quite important when thinking about the semantics of ordinary English set theory, I don't think it has much to do with the issues underlying Skolem's Paradox. After all, if we simply reformulate the paradox in terms of some particular explication of "x is countable" — e.g., by rewriting claim 2 more explicitly — then we can avoid coding issues altogether. For this reason, I'll largely bypass these issues here (see [Bays, 2000], 1.2.1–1.2.2 for more on the matter).

[5] Note that I'm going to resist the idea that any reasonably attractive version of Skolem's Paradox can be formulated in terms of an *arbitrary* countable model of ZFC. Whatever plausibility attaches to such formulations stems, I think, from some surreptitious slide between the two interpretations of "$\hat{m}$ is countable" mentioned above.

Second, the two interpretations share the same conception of *countability.* On both interpretations, claim 2 asserts the existence of a bijection between ω and some particular set, and, on both interpretations, the existence of this bijection is an issue of ordinary (naive) set theory. The difference between the two interpretations concerns the appropriate *range* of this bijection: interpretation I takes the range to be $\{x \mid x \in \hat{m}\}$, while interpretation II takes it to be $\{x \mid \mathbb{M} \models x \in \hat{m}\}$. To put this point another way, the two interpretations agree on *how* we measure the countability of a given set, but they disagree on *which* set we want to measure — i.e., which set contains the relevant "members" of the element $\hat{m}$.[6]

Given this, which of these two interpretations provides the best reading of claim 2? From one perspective, interpretation I is clearly the most natural reading of the phrase "$\hat{m}$ is countable." Further, and as we'll see later, it's the reading which makes our explication of claim 2 line up most cleanly with the syntax of $\Omega(x)$. Nevertheless, there are (at least) two difficulties with adopting interpretation I in the context of thinking about Skolem's Paradox.

First, interpretation I runs the risk of making claim 2 straightforwardly false. As Paul Benacerraf has noted, there is absolutely no reason to think that the countability of a model $\mathbb{M}$ entails that every *member* of $\mathbb{M}$ is also countable — i.e., "countable" in the sense of interpretation I.[7] In fact, it's quite easy to construct models of ZFC where the models themselves are countable but where some members of those models are uncountable (again, "uncountable" in the sense of interpretation I).

Since constructing such models lets me introduce some machinery which will eventually prove useful, I give two examples of this phenomenon here (the reader who's simply looking for the big picture should feel free to skip over these examples for the present). First, suppose that κ is an inaccessible cardinal and that $\mathbb{N}$ is a countable, elementary submodel of V_κ. In this case, even though $\mathbb{N}$ is countable, and even though $\mathbb{N} \models \text{ZFC}$, $\mathbb{N}$ still contains the uncountable set $(\aleph_1)^V$ as a *member*.[8] Second, suppose that $\mathbb{N}$ is *any* countable model for ZFC and that X

---

[6] It's important to keep this particular how/what distinction in mind. As we move along, we'll encounter some formulations of Skolem's Paradox which turn on reinterpreting the notion of countability — i.e., on changing *how* we assess the countability of some fixed set. We'll encounter other formulations which turn on varying the set whose countability we wish to assess — i.e., on changing *which set* is supposed to be uncountable. Keeping these issues distinct, therefore, will be important for understanding the mathematical issues underlying the various formulations.

[7] See [Benacerraf, 1985], 102–3.

[8] For our purposes, there are two things which are important about this example. First, the fact that κ is inaccessible entails that the model $\langle V_\kappa, \in\rangle$ satisfies ZFC. Second, the fact that $\mathbb{N}$ is an elementary submodel of V_κ entails both that $\mathbb{N}$ also satisfies ZFC and that $\mathbb{N}$ and V_κ agree on the identity of cardinals which have unique first-order definitions — e.g., cardinals like $\aleph_1$, $\aleph_2$ and $\aleph_\omega$. Hence, each of these (uncountable) cardinals must be an actual *member* of $\mathbb{N}$. As a result, the countable model $\mathbb{N}$ is literally bursting with uncountable elements.

is any set which doesn't happen to be a member of the domain of $\mathbb{N}$. Then, by simply substituting X for some arbitrary member of $\mathbb{N}$ and then modifying the "membership" relation on $\mathbb{N}$ so as to respect this substitution, we obtain another model $\mathbb{N}'$ which 1.) contains X, 2.) has exactly the same cardinality as $\mathbb{N}$, and 3.) satisfies exactly the same sentences as $\mathbb{N}$ (e.g., ZFC). If, therefore, X happens to be an uncountable set, then $\mathbb{N}'$ will be a countable model of ZFC which contains an uncountable set as a member.[9]

This, then, is one problem with taking interpretation I as the appropriate reading of "$\hat{m}$ is countable" in claim 2. Fortunately, this problem isn't as serious as it may appear to be at first. If we exercise a little care in choosing our model $\mathbb{M}$, then we can ensure that $\hat{m}$ *really is* countable in (even) the interpretation I sense. So, for instance, Paul Benacerraf has suggested that we reformulate Skolem's Paradox in terms of *transitive* models.[10] If we do so, then we ensure that for every $m \in \mathbb{M}$, $\{x \mid x \in m\} \subset M$ — in fact, we ensure that $\{x \mid x \in m\} = \{x \mid \mathbb{M} \models x \in m\}$.[11] Hence, if $\mathbb{M}$ itself is countable, then so is $\{x \mid x \in \hat{m}\}$, and Benacerraf's problem simply vanishes.

Unfortunately, transitive models are sometimes hard to come by. If we assume the existence of an inaccessible cardinal — as we did, for instance, in the first example of the second-to-last paragraph — then we can obtain such models easily.[12] Without such an assumption, however, transitive models may be hard to find. It is consistent with ZFC, for example, to accept the existence of non-transitive models

---

[9] This second example uses a technical trick which will reappear frequently throughout this paper, so it is useful to take a few moments and explain it in more detail. The example depends on two theorems of model theory. First, if two models $\mathbb{N}$ and $\mathbb{N}'$ are *isomorphic* — i.e., if there exists a bijection $f : N \to N'$ such that for every $a, b \in N$, $a \in_{\mathbb{N}} b \iff f(a) \in_{\mathbb{N}'} f(b)$ — then these models must also be *elementarily equivalent* — i.e., for every sentence ϕ, $\mathbb{N} \models \phi \iff \mathbb{N}' \models \phi$.

Second, if $\mathbb{N}$ is a model and if $f : N \to A$ is a bijection, then f carries with it a canonical method for building a model which has A as its domain and which is isomorphic to $\mathbb{N}$. To obtain this model, we simply define a relation $\in_A$ on $A \times A$ as follows:

$$a \in_A a' \iff f^{-1}(a) \in_{\mathbb{N}} f^{-1}(a').$$

Given this definition, $\langle A, \in_A \rangle$ is the desired model, and f itself is the desired isomorphism.

Returning to the example from the text, we find that substituting X for an arbitrary $\hat{n} \in \mathbb{N}$ amounts to constructing a bijection $f : N \to (N \cup \{X\}) \setminus \{\hat{n}\}$ such that $f(\hat{n}) = X$ and $f \upharpoonright (N \setminus \{\hat{n}\}) = \mathrm{Id}$. Similarly, redefining $\in$ in the manner suggested above amounts to building the very model which this bijection canonically induces. Given this, claim 1 follows directly from the definition of $\mathbb{N}'$, claim 2 follows from the fact that f is a bijection, and claim 3 follows from the fact that f is an isomorphism along with the fact that isomorphic models are elementarily equivalent.

[10] See [Benacerraf, 1985], 102–3. I will discuss transitive models in some detail when we get to section 3, so I won't say much about them here. For present purposes, the fact mentioned in the main text — i.e., that $\{x \mid x \in m\} \subset M$ when $\mathbb{M}$ is transitive — is enough to be going on with.

[11] So, interpretations I and II *coincide* for transitive $\mathbb{M}$.

[12] The technique for obtaining such models involves a result called the "Mostowski Collapsing Lemma." This lemma allows us to take any well-founded model — i.e., any model which contains no infinite descending $\in$-chains — and find a transitive model which is isomorphic to it. Hence, if we start with an inaccessible cardinal κ and then apply the Collapsing Lemma to some countable, elementary submodel of V_κ, we end up with a countable, transitive model of ZFC (see footnotes 8 and 9 for further background concerning this construction).

of set theory while rejecting the existence of transitive ones.[13] Indeed, it is fairly easy to find consistent extensions of ZFC which are incompatible with transitivity: even if ZFC has transitive models, these extensions do not.[14]

This, then, brings me to a second technique for solving Benacerraf's problem — i.e., for building our model $\mathbb{M}$ so as to ensure that $\{x \mid x \in \hat{m}\}$ is really countable. Let $\mathbb{N}$ be an *arbitrary* model of ZFC and let A be a collection of countable sets such that $|A| = |N|$.[15] Employing a trick from footnote 9, we can turn A into a model for the language of set theory which is isomorphic to our original $\mathbb{N}$. This gives us a model which 1.) satisfies ZFC, 2.) has the same size as $\mathbb{N}$, and 3.) contains only countable sets as members.[16] So, if our original $\mathbb{N}$ was countable, then this new model will have exactly the properties needed to solve Benacerraf's problem — i.e., for any $m \in \mathbb{M}$, $\{x \mid x \in m\}$ will be countable.

This gives us two ways of responding to the first problem with interpretation I — to the worry that this interpretation might make claim 2 straightforwardly false. Unfortunately, the second of these responses also serves to highlight a second *problem* with interpretation I. Suppose that the model $\mathbb{N}$ from the last paragraph is *uncountable.* Then the argument of that paragraph allows us to generate a model $\mathbb{N}'$ such that 1.) $\mathbb{N}'$ satisfies ZFC, 2.) $\mathbb{N}'$ has the same size as $\mathbb{N}$ (indeed $\mathbb{N}'$

---

[13] Here, I use the fact that if $\mathbb{M}$ is a transitive model of ZFC, and if $\mathbb{M} \models \exists \mathbb{N}$ "$\mathbb{N}$ is a transitive model of ZFC," then $\mathbb{M}$ must really contain some transitive model of ZFC (to use the jargon, the property "being a transitive model of ZFC" is *absolute* between $\mathbb{M}$ and V). I also use the fact that every transitive model of ZFC satisfies the sentence $\exists \mathbb{N}$ "$\mathbb{N}$ is a model of ZFC" (since this sentence is essentially arithmetical, and transitive models get arithmetical sentences right).

Suppose, then, that there *is* a transitive model of ZFC. As an infinite descending sequence of transitive models violates the axiom of foundation, there must be a transitive model which contains no other transitive models as members (a so-called *minimal* transitive model). This model satisfies ZFC *plus* $\exists \mathbb{N}$ "$\mathbb{N}$ is a model of ZFC" *plus* $\neg \exists \mathbb{N}$ "$\mathbb{N}$ is a transitive model of ZFC". Hence, even if transitive models exist, it is consistent with ZFC + $\exists \mathbb{N}$ "$\mathbb{N}$ is a model of ZFC" to assume that they *don't.*

[14] Here are two ways to obtain such extensions. The most straightforward way involves adding a new constant c to our language and then adding the sentences "c is a natural number," "$c \neq 1$," "$c \neq 2$," etc. to the axioms of ZFC. The resulting theory is (by compactness) consistent; but, since the constant c names a nonstandard natural number, the theory cannot have transitive (or even well-founded) models.

Alternately, we could let T *any* consistent, axiomatizable extension of ZFC and then note that the theory $T' = T \cup \neg \operatorname{Con}(T)$ is still consistent but fails to have transitive models (since, in any model of T', the "natural number" witnessing $\neg \operatorname{Con}(T)$ has to be nonstandard).

[15] Let me introduce some machinery here. Our goal is to find a set A such that 1.) A has the same size as N and 2.) every member of A is a countably infinite set. Let $\mathcal{P}_{\omega_1}(N)$ be an abbreviation for $\{X \mid X \subset N \text{ and } |X| < \omega_1\}$. Because N is infinite, we know that there are at least $|N|$ many countable subsets of N. Hence, we can find a *subset* of $\mathcal{P}_{\omega_1}(N)$ which has the same size as N. This subset gives us just the A we want.

[16] It is worth noting that there is nothing special about the fact that our final model contains only *countable* sets. The same technique can be used to obtain a model all of whose members are *finite,* and a minor modification let us obtain a model all of whose members have cardinality κ, for κ an arbitrary cardinal. In the first case, we let the domain of our model be a subset of $\mathcal{P}_{\omega}(N)$ rather than $\mathcal{P}_{\omega_1}(N)$; in the second, we let this domain be a subset of $\mathcal{P}_{\kappa^+}(\kappa) \setminus \mathcal{P}_{\kappa}(\kappa)$. (Note that we use κ rather than N in this construction, because $\mathcal{P}_{\kappa^+}(N) \setminus \mathcal{P}_{\kappa}(N)$ may be empty if $|N| < \kappa$).

is isomorphic to $\mathbb{N}$), and 3.) $\mathbb{N}'$ contains only countable sets as members. Given this, and given that we're taking interpretation I as our reading of claim 2, we can clearly use $\mathbb{N}'$ as the basis for a new version of Skolem's Paradox.[17]

But surely something's gone wrong here. Skolem's Paradox is supposed to involve the fact that *countable* models of set theory satisfy sentences like "$\hat{m}$ is uncountable." We now have a version of the paradox which uses only the *uncountable* model $\mathbb{N}'$. Indeed, since any model of set theory — whether countable or uncountable — is isomorphic to a model all of whose members are countable, we can generate versions of "Skolem's Paradox" for models of any size — and, indeed, any isomorphism type — we happen to want.

There's a flip side to this problem. Not only does interpretation I make the size of our model irrelevant, it also makes the sentence "x is uncountable" irrelevant. Once again, let $\mathbb{N}$ be an arbitrary model of ZFC. Applying tricks from the last few paragraphs, we can find a model $\mathbb{N}'$ such that 1.) $\mathbb{N}'$ is isomorphic to $\mathbb{N}$ and 2.) $\mathbb{N}'$ has only singletons as members.[18] Then, if we give "interpretation I" style readings to phrases like "is the empty set," "is a doubleton," "is infinite," etc., we can generate obvious analogs of Skolem's Paradox for those phrases.[19]

Together, these examples show that there is something *conceptually* wrong with using interpretation I to make sense of Skolem's Paradox. A proponent of Skolem's Paradox thinks that there is something puzzling about the fact that countable models of set theory can satisfy sentences like "$\hat{m}$ is uncountable." On interpretation I, the fact that these models are countable is irrelevant, and the puzzle at issue can be formulated for sentences which are *far* simpler than those involving countability/uncountability. As a result, interpretation I seems to miss the *point* of Skolem's Paradox.

This brings me to interpretation II. Clearly, interpretation II avoids the two problems we've just been discussing. If $\mathbb{M}$ is countable, then every set of the form $\{x \mid \mathbb{M} \models x \in m\}$ is also countable; hence, Benacerraf's problem doesn't arise. Further, it's *because* $\mathbb{M}$ is countable, that $\{x \mid \mathbb{M} \models x \in m\}$ has to be countable; so, the countability of $\mathbb{M}$ plays, as it should, a real role in our argument.[20] Nor does the argument generalize to arbitrary set-theoretic properties.

---

[17] That is, since $\mathbb{N}' \models$ ZFC, there must be some $\hat{n} \in \mathbb{N}'$ such that $\mathbb{N}' \models \Omega[\hat{n}]$. By our construction, however, every member of $\mathbb{N}'$ is countable (in the interpretation I sense of the phrase). So, we get obvious analogs of claims 1 and 2 above.

[18] We might, for instance, let $A = \{\{n\} \mid n \in \mathbb{N}\}$ and then follow through the argument from the third-to-last paragraph.

[19] For example, let $\Omega'(x)$ be the formula which "codes up" the phrase "x is the empty set." Since $\mathbb{N}' \models ZFC$, there must be some $\hat{n} \in \mathbb{N}'$ such that $\mathbb{N}' \models \Omega'[\hat{n}]$. Clearly, however, $\hat{n}$ isn't *really* empty; by construction, $\hat{n}$ is *really* a singleton.

Note that this argument is perfectly general. If P is a set-theoretic property such that there are infinitely many sets which *don't* have P, then our isomorphism trick lets us build a model, $\mathbb{N}$, such that no member of $\mathbb{N}$ has P. So, if ZFC $\vdash \exists x\, P(x)$, then we can generate an "interpretation I"-style analog of Skolem's Paradox for the property P.

[20] Some cautions are in order here. With enough care, it's possible to build uncountable models which exhibit Skolem's Paradox-like phenomena (we'll see some in sections 4–5). Nonetheless, interpretation II does two things for us. It ensures that *every* countable model gives rise to a version of Skolem's Paradox, and it ensures that uncountable models need to have a special

If $\mathbb{M} \models$ "$\hat{m}$ is the empty set," then $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ really is the empty set; if $\mathbb{M} \models$ "$\hat{m}$ is a doubleton," then $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ really is a doubleton; etc.[21] Hence, interpretation II does a better job of capturing the point of Skolem's Paradox than interpretation I did.

That being said, interpretation II does have one, relatively minor, problem. If we use interpretation II as the basis for explicating claim 2, then it's not obvious that our explication will line up syntactically with the $\Omega(x)$ in claim 1. On the surface, explicating the claim "$\{x \mid \mathbb{M} \models x \in \hat{m}\}$ is countable" should involve a fair bit of machinery that's devoted to characterizing the model $\mathbb{M}$ and to cashing out the notion of satisfaction. But, there's nothing corresponding to this machinery in (the most natural version of) the formula $\Omega(x)$. On the purely syntactic level, it's the explication of "$\{x \mid x \in \hat{m}\}$ is countable" which lines up most cleanly with the formula $\Omega(x)$.

Fortunately, there are several ways of overcoming this problem. First, we could choose our model $\mathbb{M}$ so as to ensure that interpretations I and II *agree* on this model. If we let $\mathbb{M}$ be transitive, for instance, then $\{x \mid \mathbb{M} \models x \in m\} = \{x \mid x \in m\}$ for every $m \in \mathbb{M}$. Similarly, if we start with a countable $\mathbb{M}$ and an arbitrary $\hat{m} \in \mathbb{M}$, then a simple variant of our footnote 9 trick will allow us to find an isomorphic $\mathbb{M}'$ and $\hat{m}'$ such that $\{x \mid \mathbb{M}' \models x \in \hat{m}'\} = \{x \mid x \in \hat{m}'\}$.[22] In either of these cases, then, the problem from the last paragraph disappears: for *these* models, the syntax of $\Omega(x)$ lines up with a perfectly natural explication of "$\{x \mid \mathbb{M} \models x \in \hat{m}\}$ is countable."

Second, since we're particularly interested in the membership relation *on* $\hat{m}$, we could simply use a new symbol to represent this relation. So, for instance, let $\mathbb{M}$ be a countable model, let $\hat{m}$ be arbitrary element of $\mathbb{M}$, and let "$\in_{\hat{m}}$" be a new binary relation. Expand $\mathbb{M}$ so as make $\in_{\hat{m}}$ represent "membership" in $\hat{m}$.[23] Then, there's a natural formula $\Omega'(x)$ such that $\mathbb{M}' \models \Omega'[\hat{m}]$ and such that the syntax of $\Omega'(x)$ lines up cleanly with an equally natural explication of "$\{x \mid \mathbb{M}' \models x \in_{\hat{m}} \hat{m}\}$ is countable."[24] This gives us a second technique for making interpretation II

---

isomorphism-type if they are to give rise to Skolem's Paradox (so, it's not the case that *every* model of ZFC is isomorphic to a model in which a variant of Skolem's Paradox arises).

[21] Of course, there will still be notions other than countablity/uncountablity which the model gets wrong — e.g., "x is finite," "x is inaccessible," "x is the power set of y," etc. But these are relatively complicated set-theoretic notions, so it's not too surprising that models which get countablity/uncountablity wrong should also have problems with them. What interpretation II does is to ensure that this problem isn't *completely general;* on interpretation II, our models get *easy* notions — "being empty," "being a singleton,"etc. — correct.

[22] If $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ isn't already a member of the domain of $\mathbb{M}$, then we can just replace $\hat{m}$ with $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ to get our $\mathbb{M}'$. If $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ *is* a member of $\mathbb{M}$, then we can let a be any set which isn't a member of $\mathbb{M}$. We get $\mathbb{M}'$ by first replacing $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ with a, and then replacing $\hat{m}$ with $\{x \mid \mathbb{M} \models x \in \hat{m}\}$.

[23] Some clarification may be in order here. In expanding $\mathbb{M}$ we're not *adding* anything to $\mathbb{M}$'s domain — indeed, we're not *changing* $\mathbb{M}$'s domain at all. Nor are we changing the way $\mathbb{M}$ interprets the symbol "$\in$." We're simply stipulating that the expanded model, $\mathbb{M}'$, *also* interprets the symbol $\in_{\hat{m}}$ via the clause: $\mathbb{M}' \models m_1 \in_{\hat{m}} m_2 \Longleftrightarrow m_2 = \hat{m}$ and $\mathbb{M}' \models m_1 \in m_2$.

[24] The formula $\Omega'(x)$ is obtained by taking our original $\Omega(x)$ and replacing each instance of $y \in x$ with $y \in_{\hat{m}} x$. The explication uses $\in_{\hat{m}}$ as an abbreviation for $\mathbb{M}' \models x \in \hat{m}$. Note that,

work. Unlike the first, it allows us to start with an arbitrary countable model of ZFC; but, like the first, it still requires us to use some trickery to make the $\Omega(x)$ in claim 1 line up with a natural explication of claim 2.

In the long run, though, this kind of trickery is probably unavoidable. The preceding discussion shows that, if we want to make Skolem's Paradox look plausible, then we need to find an interpretation of claim 2 which satisfies the following three conditions: 1.) it makes claim 2 come out true, 2.) it ensures that the truth of claim 2 is appropriately *connected* to the fact that $\mathbb{M}$ is a countable model of ZFC, and 3.) it ensures that the syntax of our explication of claim 2 lines up neatly with the syntax of $\Omega(x)$. Interpretation I does a good job with condition 3, but it requires some tricks to deal with condition 1 and it can't deal with condition 2 at all. Interpretation II takes care of conditions 1 and 2, but it requires some tricks to take care of condition 3. In both cases, therefore, we need some tricks to ensure that our three conditions are jointly satisfied — in particular, we need some constraints on the choice of our model $\mathbb{M}$.

This need for care in choosing $\mathbb{M}$ brings me to my third preliminary point. So far, our discussion has pretty much ignored our initial stipulation that $\mathbb{M} \models \text{ZFC}$. (We've only used it to ensure that there exists some $\hat{m} \in \mathbb{M}$ such that $\mathbb{M} \models \Omega[\hat{m}]$.) Clearly, though, the fact that $\mathbb{M} \models \text{ZFC}$ plays a larger role in making Skolem's Paradox look plausible. After all, it's not the *members* of $\mathbb{M}$ which make us think that this model has something to do with set theory: there are many models for the language of set theory which contain objects other than sets, and there are some models which contain no sets at all.[25] So, unless these models satisfy some set-theoretic axioms — say, a significant fragment of ZFC — it's hard to see why they should be regarded as having *anything* to do with our topic.

To reinforce this point, we should notice *just how badly* models for the language of set theory can fail to satisfy ZFC, while nevertheless satisfying formulas like $\Omega(x)$. Consider the model whose domain consists of the numbers 1–10 and which interprets "$\in$" by:

$$n \in m \Longleftrightarrow n \leq 5 \text{ and } 5 < m \leq 10.$$

In this model, all numbers greater than 5 satisfy $\Omega(x)$, although the model itself has no connection with set theory and fails to satisfy even the axiom of extensionality.[26]

---

because $\mathbb{M}' \models m \in_{\hat{m}} \hat{m} \Longleftrightarrow \mathbb{M}' \models m \in \hat{m}$, this also serves as a reasonable explication of "$\{x \mid \mathbb{M}' \models x \in \hat{m}\}$ is countable."

[25] We might, for instance, build a model which contained only my three cats as elements and which interpreted "$\in$" as identity. This model wouldn't be very interesting — and it certainly wouldn't satisfy the axioms of set theory — but it would be a model for the *language* of set theory.

[26] With respect to the axiom of extensionality, note that all of the numbers $n \leq 5$ have exactly the same "members," as do all of the numbers $m > 5$. With respect to the satisfaction of $\Omega(x)$, note that this formula has the overall form:

$$\Omega(x) \equiv_{df} \neg\exists f \text{ [``}f \text{ is a bijection \& Domain}(f) = \omega \text{ \& Range}(f) = x].$$

Here, the phrases "x is a bijection," "Range(f)," and "Domain(f)" are themselves mere abbreviations for further (rather complicated) formulas. For our purposes, the important thing to notice is that the formulas "f is a bijection and "Range$(f) = x$" together entail that every member of

For that matter, if we let $\Psi(y)$ be the formula which codes "$y = \omega$," then *any* model which satisfies "$\neg\exists y\,\Psi(y)$" will also satisfy "$\forall x\,\Omega(x)$."[27] So, unless we're working with a model which satisfies some basic set-theoretic axioms, there's just no reason to think that the formula $\Omega(x)$ has any special significance.

At this point, then, we have an overview of the machinery needed to set up Skolem's Paradox and to make it look somewhat plausible. We start with a countable model for the language of set theory, $\mathbb{M}$. This model has several nice properties. Most importantly, $\mathbb{M} \models \text{ZFC}$; but $\mathbb{M}$ also satisfies one of the structural constraints discussed on pages 618–624 (e.g., $\mathbb{M}$ is transitive, or it's been expanded with an appropriate $\in_{\hat{m}}$ relation, or ...). Next, we note that there's a formula $\Omega(x)$ — a formula which it's awfully hard to resist abbreviating with the phrase "x is uncountable" — and an element $\hat{m} \in \mathbb{M}$ such that $\mathbb{M} \models \Omega[\hat{m}]$. This gives us, once again, the two claims highlighted on page 617:

1. $\mathbb{M} \models \Omega[\hat{m}]$
2. $\hat{m}$ is countable.

Finally, we provide a natural explication of the phrase "$\hat{m}$ is countable" in claim 2 which 1.) follows the lead given by interpretation II from page 618 and 2.) uses no symbols other than $=, \in, \neg, \rightarrow$, and $\exists y$ (and, perhaps, $\in_{\hat{m}}$ and/or some punctuation).

Given all this, Skolem's paradox arises from two things. First, the sentence produced by our explication of "$\hat{m}$ is countable" is true. (Since $\mathbb{M}$ is countable, $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ is also countable, and our sentence is just a longwinded way of *saying that* $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ is countable.) Second, this sentence *looks like* an unnegated version of the formula $\Omega(x)$. That is, if we simply inspect the syntax of these two expressions — ignoring the initial negation in $\Omega(x)$ — then we will find that they contain *exactly the same symbols* in *exactly the same order*. Together, these two facts explain why claims 1 and 2 may still look quite problematic: both of the claims are true, and the formula that $\mathbb{M}$ satisfies in claim 1 looks just like the negation of claim 2 (after, of course, claim 2 has been appropriately explicated).

That being said, looks aren't everything, and syntax isn't semantics. To make Skolem's Paradox *work* — as opposed to simply making it look superficially plausible — we need to uncover a stronger connection between $\Omega(x)$ and some particular explication of "x is uncountable." Ideally, we would like to find a deep semantic connection between the two expressions: perhaps they *mean the same*

---

x is also a member of a member of a member of f. Hence, since the interpretation of "$\in$" in our model does not allow membership chains containing more than two elements, no f of the type forbidden by, e.g., $\Omega[6]$ lives in our model. Hence, the model satisfies $\Omega[6]$ (and $\Omega[7]$, and $\Omega[8]$, etc.).

It is worth noting that this model *also* satisfies $\Omega[n]$ for $n \leq 5$, though unpacking the relevant definitions is more time-consuming in these cases and depends on a particular definition of ω. The basic idea is that discussed in the next footnote.

[27] Again, this is a simple consequence of the definition of $\Omega(x)$. To see this, simply note that:

$$\neg\exists y\,\Psi(y) \vdash \neg\exists f\ [\cdots\ \&\ \exists y\ (y = \text{Domain}(f)\ \&\ \Psi(y))\ \&\ \cdots]$$

for *any* possible values of "$\cdots$" (including those relevant to $\Omega(x)$).

thing or *have the same sense.* At the very least, we need to establish a truth-functional implication between the formula $\Omega(x)$, *as this formula gets interpreted at the model* $\mathbb{M}$, and the particular explication in question. Without such a connection, Skolem's Paradox won't get off the ground.

For convenience in discussing these issues, let me introduce two pieces of notation. First, I will use $\Omega_E(x)$ to denote our canonical explication of "x is uncountable."[28] Second, I will use $\Omega_{\mathbb{M}}(x)$ to denote the interpretation of the formula $\Omega(x)$ on the model $\mathbb{M}$. That is, $\Omega_{\mathbb{M}}(x)$ is the interpretation of $\Omega(x)$ which results from letting the quantifiers in $\Omega(x)$ range over the domain of $\mathbb{M}$, letting the significance of $\in$ (and, perhaps, $\in_{\hat{m}}$) be fixed by the interpretation function of $\mathbb{M}$, and letting the significance of $\neg, \rightarrow$, and $=$ be given by the recursion clauses in the the definition of first-order satisfaction. With these abbreviations in place, the above discussion shows that Skolem's Paradox turns on some variant of the following claim:

$$\forall m \in \mathbb{M}\,[\Omega_{\mathbb{M}}(m) \implies \Omega_E(m)]. \tag{\dagger}$$

This claim captures — in a purely truth-functional manner — the kind of connection between $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ which would have to hold if Skolem's Paradox were to constitute a genuine mathematical contradiction.[29] To solve the paradox, therefore, we simply need to figure out what's wrong with (†).

Of course, from one perspective, it's easy to see what's wrong with (†): *it's false.* On the one hand, if (†) *were* true, then we could use Skolem's Paradox itself to generate a straightforward contradiction in set theory. Since set theory *isn't* contradictory, we should obviously apply *modus tollens* and reject claims like (†). On the other hand, it's relatively easy to construct models of ZFC in which, for certain elements m, $\Omega_{\mathbb{M}}(m)$ is true but $\Omega_E(m)$ is *clearly* false. (We will, in fact, construct several such models later in this paper.) This makes it look as though Skolem's Paradox can — and perhaps even *should* — be dismissed rather quickly.

---

[28] Recall, here, that $\Omega_E(x)$ is not generated by *interpreting* a formula of first-order set theory. We do not, that is, *begin* with a string of uninterpreted first-order symbols and then *stipulate* that these symbols are to be understood in some particular way. Instead, we begin with a sentence of ordinary mathematical English, and then use a certain collection of symbols — which just happen to be commonly used in the formulation of first-order set theory — as abbreviations for terms and phrases which already occur in this sentence. As a result, $\Omega_E(x)$ has exactly the same semantics as an ordinary language explication of "x is uncountable" — i.e., a completely unabbreviated one.

Of course, the fact that $\Omega_E(x)$ has the same semantics as this ordinary English explication doesn't mean that $\Omega_E(x)$ is semantically unproblematic. If there are problems with our original explication of "x is uncountable" — e.g., problems of vagueness or ambiguity — then these will carry over to $\Omega_E(x)$. It does, however, mean that there are no *special* problems (which arise merely) from the fact that $\Omega_E(x)$ makes (purely abbreviatory) use of the symbols $=, \in, \neg, \rightarrow$ and $\exists$.

[29] In [Bays, 2000], I isolate a general form of argument which encompasses many different versions of Skolem's Paradox. I show that (†) provides a necessary condition for any argument of this form to be sound, and I show that (†) provides a sufficient condition for at least one argument of this form to be sound. In this sense, then, (†) really does lie at the heart of (the mathematical side of) Skolem's Paradox. For reasons of space, I'll eshew a full discussion of these different variants of Skolem's Paradox here. For more on the subject, see chapter 1 of [Bays, 2000] (especially section 1.2.2).

Once again, though, I think Skolem's Paradox is a bit harder than this. For one thing, although the above argument shows *that* (†) is false, it doesn't really explain *why* it's false. That is, it doesn't provide an analysis of the semantic differences between $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ which explains why the former does not entail the latter (or, at the very least, why the semantics of the two are sufficiently different that we should not be *surprised* when the former doesn't entail the latter with respect to a particular model $\mathbb{M}$).

For another thing, this approach may seem to miss the *point* of Skolem's Paradox. Someone worried about Skolem's Paradox *starts out* thinking that there's enough of a relationship between $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ that we should seriously consider re-construing classical set theory in light of this relationship. That is, he is at least tempted by the idea that Skolem's Paradox shows that classical set theory, when taken at face value, *just is* contradictory, and that we need to appeal to philosophical notions like *relativity* or *perspective* to ease the sting of this contradiction.

Given this, I think it is highly unlikely that a proponent of Skolem's Paradox would be persuaded by the kind of *modus tollens* argument I just gave. This proponant already *knows* that assumptions like (†) lead to contradictions — that, after all, is the whole *point* of Skolem's Paradox. By themselves, however, these contradictions don't lead him to abandon (†). Hence, unless my *modus tollens* argument is supplemented by a more detailed analysis of *why* (†) fails — of *where* the semantics of $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ differ and of *how* this difference leads to the failure of (†) — the proponent of Skolem's Paradox is unlikely to find it persuasive.

2 A QUICK TECHNICAL SOLUTION

In this section, I discuss two, fairy obvious, differences between the semantics of $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$. Together, they explain why there's nothing at all surprising about the failure of claims like (†). In doing so, they bolster the plausibility of the basic *modus tollens* argument given at the end of the last section, and they show why there's no purely mathematical reason to be worried about Skolem's Paradox.

Before beginning this discussion, a philosophical comment is in order. The solution to Skolem's Paradox that I sketch here — a solution I call the "technical solution" — simply explains why there's no straightforward contradiction between naive set theory and the Löwenheim-Skolem theorem. With a little care, it can also be used to explain why the Löwenheim-Skolem theorem doesn't introduce contradictions into various forms of axiomitized set theory. As a result, the technical solution allows the working set theorist — or the philosopher who is content to take a naively realistic attitude toward set theory — to remain untroubled by Skolem's Paradox.

Of course, many philosophers will be reluctant to take such an attitude toward set theory — e.g., those with *theoretical* reasons for identifying the semantics of $\Omega_E(x)$ with those of $\Omega_{\mathbb{M}}(x)$, or even just those who have qualms about the overly-quick invocation of things like "the ordinary English significance of '$\in$'." Such philosophers are unlikely to find the solution developed in this section satisfactory. However, because the main topic this paper — the mathematical side of Skolem's Paradox — has more to do with the fine *details* of the technical solution than with its ultimate philosophical adequacy, I won't say too much about these philosophers' worries here (just a little bit in section 6). For more on their concerns, see [Bays, 2000] (chapter 3), [Benacerraf, 1985], [Hart, 1970], [Klenk, 1976], [Myhill, 1967], or [Resnik, 1966].[30]

What, then, can we say about the semantic differences between $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$? First, we can note that the semantics of $\Omega_E(x)$ interpret the symbol "$\in$" so that:

$E_\in$: "$x \in y$" is true iff y is a set and x is a member of y.

In contrast, let $i_{\mathbb{M}}$ be the interpretation function for $\mathbb{M}$.

Then the semantics of $\Omega_{\mathbb{M}}(x)$ interpret "$\in$" so that

$\mathbb{M}_\in$: "$x \in y$" is true iff $\langle x, y\rangle$ is a member of $i_M(\in)$.

Clearly, however, there is *no* reason to think that these two interpretations of "$\in$" are coextensive. This is most obvious when some elements of $\mathbb{M}$ aren't even genuine *candidates* for the ordinary membership relation. It is possible, for instance, to build models of ZFC in which the "membership relation" holds between ordinary housecats.[31] Similarly, providing that there are infinitely many non-sets in the world, we can find models of ZFC whose domains contain no sets at all.[32] In cases like these, it should be quite clear that the semantics of $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ are interpreting the symbol "$\in$" in radically different ways.

Further, even when a model *does* contain sets — and perhaps even *only sets* — there is no guarantee that this model's interpretation of "$\in$" agrees with the ordinary English interpretation of this symbol. To illustrate this point, let $\mathbb{N}$ be an arbitrary model of ZFC, let X be the collection of singletons of members of $\mathbb{N}$, and let Y be the collection of doubletons of members of X. Applying our trick from footnote 9, we can build a model $\mathbb{N}'$ such that 1.) $\mathbb{N}'$ has Y as its domain and 2.) $\mathbb{N}'$ is isomorphic to $\mathbb{N}$ (and, hence, satisfies exactly the same sentences as $\mathbb{N}$ does). Given this construction, all of the members of $\mathbb{N}'$ are genuine sets, but $\mathbb{N}'$ displays almost *no* agreement with ordinary English concerning the interpretation of "$\in$." In particular, there are *many* sets $n_1, n_2 \in \mathbb{N}'$ such that $\mathbb{N}' \models n_1 \in n_2$, but there are *no* sets $n_1, n_2 \in \mathbb{N}'$ such that $n_1 \in n_2$.[33]

---

[30] I should probably also note that I don't view the technical solution as in any way *original.* Others have said quite similar things (see, e.g., [Benacerraf, 1985], [Myhill, 1967] or [Resnik, 1966]). Instead, the material in this section is preparatory for the more-detailed discussions of quantification and membership presented in sections 3–5.

[31] To build such a model, we just let $\mathbb{N}$ be an arbitrary model of ZFC, and we let n and n' be arbitrary elements of $\mathbb{N}$ such that $\mathbb{N} \models n \in n'$. Given this, let Puffy and Fluffy be two ordinary housecats (neither of which lives in the domain of $\mathbb{N}$), and let $f : N \to (N \setminus \{n, n'\}) \cup \{\text{Puffy}, \text{Fluffy}\}$ be a bijection such that $f(n) = \text{Puffy}, f(n') = \text{Fluffy}$ and $f \restriction (N \setminus \{n, n'\}) = \text{Id}$. Then, using f to induce a canonical "membership" relation on the domain $(N \setminus \{n, n'\}) \cup \{\text{Puffy}, \text{Fluffy}\}$ — i.e., in the manner described in footnote 9 — we obtain a model $\mathbb{N}'$ such that $\mathbb{N}' \models \text{ZFC} +$ "Puffy $\in$ Fluffy."

[32] Again, this follows from a simple application of our trick from footnote 9. We start by letting $\mathbb{N}$ be an arbitrary countable model of ZFC. We then let X be a countable collection of non-sets, and let $f : N \to X$ be some arbitrary bijection. Following the argument of footnote 9, we note that f induces a relation, $\in_f$, on X such that the model $\langle X, \in_f\rangle$ is isomorphic to $\mathbb{N}$. Hence, $\langle X, \in_f\rangle$ satisfies ZFC as desired.

These examples show that the semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ sometimes disagree about expressions of the form "$a \in b$." When we move to more complicated expressions, we find further disagreements. In particular, the semantics of $\Omega_E(x)$ interpret the expression "$\exists\, x$" as synonymous with the phrase "there is a set x, such that" (since the former is, after all, simply an *abbreviation* for the latter). In contrast, the semantics of $\Omega_{\mathbb{M}}(x)$ interpret "$\exists\, x$" via the recursion clause:

$$\exists. \qquad \mathbb{M} \models \exists\, x\, \Phi(x) \iff \text{there exists an } m \in M \text{ such that } \mathbb{M} \models \Phi[m].$$

In practice, this amounts to identifying the expression "$\exists\, x$" with the phrase "there is an element $x \in \mathbb{M}$, such that." Given that the domain of $\mathbb{M}$ is not identical with the set-theoretic universe (as $\mathbb{M}$ is, after all, a merely *countable* model), this introduces a second difference between the semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$.

Let me make a few comments concerning these two semantic differences. To begin: there shouldn't be anything surprising — from either a mathematical or a philosophical standpoint — about the fact that first-order model theory allows us to vary the interpretation of $\in$ and $\exists$ (and, as a result, that it doesn't "capture" the ordinary English notions of *membership* and *quantification over the set-theoretic universe*). From a mathematical standpoint, model theory is *designed* to allow substantial variation in the models at which particular sentences can be interpreted (and, indeed, in the models at which particular sentences can come out true). The point of model theory is to investigate the *interaction* between models and formulas. So, if we give our formulas too specific a semantics — e.g., by fixing *everything* about the interpretation of our language and leaving *nothing* to vary as we move from model to model — then we threaten to make those formulas model-theoretically trivial.[34]

In the special case of *first-order* model-theory, we fix the interpretation of $\neg$, $\rightarrow$, and $=$, but we allow the interpretation of quantifiers and of other relations — e.g., $\in$ or $\in_{\hat{m}}$ — to vary. (In particular, therefore, we don't even *try* to fix the significance of "$\in$" or "$\exists$.") The resulting system is interesting in part because has nice meta-theoretic properties — e.g., completeness and compactness — which render it easy to work with. More importantly, when we use first-order model theory to investigate the axioms of *set theory*, we find that the ability to reinterpret $\in$ and $\exists$ as we move from model to model underlies some standard set-theoretic techniques

---

[33] This latter claim follows from the fact that every element of Y *is* a doubleton which *contains* only singletons. Hence, there are no elements $y_1, y_2 \in A$, such that $y_1 \in y_2$.

[34] In particular, then, we shouldn't be surprised to find that first-order sentences can be satisfied by a whole variety of structurally different models. In designating a sentence "first-order," we say that it is to be evaluated at *these kinds* of models. And while *some* sentences — e.g., $\forall x \forall y\; x = y$ — may do a good job at picking out the structure of their models, this cannot be the case for all sentences. If it were, then first-order model theory would lose much of its mathematical interest.

— forcing, inner models, large-cardinal arguments, etc. These techniques turn out to be important for understanding the structure of the *real* set-theoretic universe. Hence, not only is there nothing surprising about the fact that first-order model theory doesn't capture the ordinary English significance of "$\in$" or "there is a set," but there's also nothing surprising about the fact that mathematicians — i.e., model theorists and set theorists — continue to study it anyway.

From a more philosophical standpoint, the fact that model theory lets us vary the interpretation of certain symbols is part of what makes the subject philosophically fruitful. It is, for instance, what allows us to give model-theoretic analyses of the notion of *logical consequence,* and it's what lets us to use models as formal proxies for possible worlds in certain metaphysical arguments. Once again, then, the fact that model theory doesn't fix the significance of every symbol in our language — e.g., "$\in$" in the case of first-order model theory — shouldn't be viewed as a surprising *flaw* in the model-theoretic machinery; on the contrary, it's part of what makes this machinery so philosophically useful. To put this point in Microsoft's jargon, variability of interpretation is a *feature* of first-order model theory, it's not a *bug.*

This brings me to two final comments. First, it's important to note that differences in the way $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret "$\in$" and "$\exists\, x$" give rise to *many* differences in the overall interpretation of these two expressions. Because each expression contains several thousand instances of "$\in$" and "$\exists\, x$," there will be *many* places where the semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ diverge. Hence, ground level differences in the interpretation of "$\in$" and "$\exists\, x$," have the potential to ramify into deeper — and far more systematic — differences between the overall semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$. So, to the extent that we find differences in the interpretation of "$\in$" and "$\exists\, x$" unsurprising, we should find differences in the overall semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ even less surprising.

Second, these differences in the semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ exist *even when* the expressions happen to agree about some particular element of $\mathbb{M}$ — i.e., even when $\Omega_E(m)$ and $\Omega_{\mathbb{M}}(m)$ both come out true (or false) for some particular m. If we look carefully, we will usually find that these sentences are true (or false) for structurally different reasons. The membership relations which make $\Omega_E(m)$ true may have nothing to do with the instances of $\mathbb{M} \models m_1 \in m_2$ which make $\Omega_{\mathbb{M}}(m)$ true, and the particular sets which make "there exists a set x such that …" true may be different from the elements of $\mathbb{M}$ which make "there exists an $m \in M$, such that …" true. As a result, even when $\Omega_E(m)$ and $\Omega_{\mathbb{M}}(m)$ do happen to agree, we should view their agreement as little more than a happy accident.

At the end of the day, then, we should not be surprised to find that that claims like (†) fail rather frequently. Since many of the corresponding parts of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ have radically different semantics — and semantics which differ in ways which directly affect the truth values of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ — we have no reason to *expect* that the two expressions will have the same truth-value. Indeed, as noted in the last two paragraphs, the differences between $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ are sufficiently severe and pervasive that it's little more than an accident when they do happen to agree. If we find a case where they agree on *all* members of a model's domain, then this agreement itself should be viewed as a surprising fact which stands in need of explanation; in cases where they *don't* so agree, we should regard their disagreement as completely ordinary and unsurprising.

This, therefore, gives us a generic — and a somewhat simpleminded — solution to Skolem's Paradox. In its simplest formulations — e.g., that presented at the beginning of section 1 — the paradox rests on a straightforward equivocation between the (superficially similar) expressions $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$. More sophisticated formulations, although they may avoid outright equivocation, must still postulate a *connection* between $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ which is strong enough to ground claims like (†). As we have just seen, however, there is absolutely no reason to believe in such a connection. At the level of individual symbols, there are clear differences in the way $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret "$\in$" and "$\exists$"; at the level of whole expressions, postulating a connection between $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ leads to immediate contradictions — i.e., when we consider elements like $\hat{m}$. Given all this, we have no reason to regard Skolem's Paradox as a genuine mathematical problem. Indeed, on reflection, it's not even a particularly *surprising* fact about the models of first-order set theory.

3 THE VIRTUES OF QUANTIFICATION

In the last section, I gave a generic solution to Skolem's Paradox. I noted that the paradox rests on conflating the ordinary English significance of "$\in$" and "$\exists$" with the significance given to these symbols by first-order model theory — i.e., when we interpret them at a particular model. I did not, however, say anything about *which instances* of "$\in$" and "$\exists$" are really crucial to Skolem's Paradox. For a generic solution, it's enough to notice that there are many places where the semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ diverge; hence, there's nothing surprising about the fact that these two expressions often have different truth-values.

In the philosophical literature, there's a widespread tradition of wanting to go a bit further than this — of wanting, that is, to pin down just which instances of "$\in$" and "$\exists$" really serve to *explain* Skolem's Paradox. And, from one perspective, it seems like we should be able to accomplish this. Consider the formula we've been calling $\Omega(x)$. Abbreviating wildly, we can represent this formula as follows:

$$\Omega(x) \equiv \neg\exists f \text{ ``} f : \omega \to x \text{ is a bijection''}$$

where ω is the standard set-theoretic representation of the natural numbers. Clearly, any interpretation of this formula will depend heavily on the significance we give to its initial existential quantifier — i.e., to the "$\exists f$" which follows the initial negation. As we have seen, however, $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret this quantifier quite differently.

Following this line of thought, let's track the relevant differences through the details of our claim, (†). On the one hand, it's easy to see that the expression $\Omega_E(\hat{m})$ means something like:

1. There is no f in the set-theoretic universe such that $f : \omega \to \{x \mid \mathbb{M} \models x \in \hat{m}\}$ is a bijection.

On the other hand, the expression $\Omega_{\mathbb{M}}(\hat{m})$ means (at best):

2. There is no $f \in \mathbb{M}$ such that $f : \omega \to \{x \mid \mathbb{M} \models x \in \hat{m}\}$ is a bijection.

Given 1 and 2, the explanation for the failure of (†) looks quite simple. Because the domain of $\mathbb{M}$ is countable, the set $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ must also be countable. Hence, there really is a bijection between ω and $\{x \mid \mathbb{M} \models x \in \hat{m}\}$, and claim 1 is simply false. In contrast, as long as all the f's which falsify 1 happen to live outside the domain of $\mathbb{M}$, claim 2 can perfectly well be true. As a result, Skolem's Paradox simply shows that countable models don't contain all the functions which live in the set-theoretic universe (no surprise there!), and that some countable models don't contain *any* functions belonging to a particular class — i.e., the class of bijections from ω to $\{x \mid \mathbb{M} \models x \in \hat{m}\}$.

Let's look at this argument from a slightly different angle. We know that the quantifiers in $\Omega_E(\hat{m})$ range over a domain which is large enough to include several (indeed $2^{\aleph_0}$!) bijections $f : \omega \to \{x \mid \mathbb{M} \models x \in \hat{m}\}$. Further, the semantics of $\Omega_E(\hat{m})$ allows it to recognize these f's *as* bijections from ω to $\{x \mid \mathbb{M} \models x \in \hat{m}\}$. So, $\Omega_E(\hat{m})$ comes out false. The idea behind the present argument is that this kind of analysis should *almost* work for $\Omega_{\mathbb{M}}(\hat{m})$ as well. *If* $\mathbb{M}$ only knew about some bijection $f : \omega \to \{x \mid \mathbb{M} \models x \in \hat{m}\}$, *then* $\mathbb{M}$ would recognize f as a bijection from ω to $\{x \mid \mathbb{M} \models x \in \hat{m}\}$. As a result, $\mathbb{M}$ would satisfy some formula of the form "$f : \omega \to \hat{m}$ is a bijection," and it would fail to satisfy $\Omega(\hat{m})$. In short, if the quantifiers in $\Omega_{\mathbb{M}}(\hat{m})$ could only know about the same functions that the quantifiers in $\Omega_E(\hat{m})$ know about, then the analysis of $\Omega_{\mathbb{M}}(\hat{m})$ would run exactly parallel to that of $\Omega_E(\hat{m})$. However, the quantifiers in $\Omega_{\mathbb{M}}(\hat{m})$ *don't* know about the same functions as the quantifiers in $\Omega_E(\hat{m})$, and this difference is what explains the failure of claims like (†).

This, then, gives us a simple — and a relatively attractive — solution to Skolem's Paradox. It's a solution which focuses on differing interpretations of the initial existential quantifier in $\Omega(x)$, and which uses these differences to explain the failure of claims like (†). It's also a rather common solution in the philosophical literature. Variants of it can be found in [Benacerraf, 1985], [Hart, 1970], [McIntosh, 1979], and [Myhill, 1967], and it's even made its way into several introductory textbooks (see, for instance, [Shoenfield, 1967] and [van Dalen, 1997]). Further, although I don't think this quantificational analysis provides a *complete* solution to Skolem's Paradox (for reasons we'll discuss in sections 4–5), I do think it gets some things deeply right.

First, the quantificational solution is right to insist that there is a difference in the way $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret the initial existential quantifier in $\Omega(x)$. More specifically, it's right to insist that there exist bijections $f : \omega \to \{x \mid \mathbb{M} \models x \in \hat{m}\}$ which 1.) live within the range of the quantifiers in $\Omega_E(x)$ (and, in so doing, help to explain why $\Omega_E(x)$ comes out false), but which 2.) live outside the range of the quantifiers in $\Omega_{\mathbb{M}}(\hat{m})$. After all, there are only countably many elements in the domain of $\mathbb{M}$, and there are $2^{\aleph_0}$ bijections between ω and $\{x \mid \mathbb{M} \models x \in \hat{m}\}$. So, at least some of these bijections (indeed $2^{\aleph_0}$ many of them!) must live outside the domain of $\mathbb{M}$. As a result, there *really is* an important difference between the class of bijections which gets "noticed" by the initial existential quantifier in $\Omega_E(x)$ and that which gets "noticed" by the initial existential quantifier in $\Omega_{\mathbb{M}}(x)$.

Second, there are *some cases* where this difference in quantifier-ranges really does explain what's going on in Skolem's Paradox. To see this — and to further bring out real virtues of the quantificational solution to Skolem's Paradox — it's worth looking at one such case in more detail. I begin with some terminology. Let's say that a model $\mathbb{N}$ is *transitive* if 1.) every member of $\mathbb{N}$ is itself a set, 2.) every member of a member of $\mathbb{N}$ is also a member of $\mathbb{N}$ and 3.) the "membership" relation on $\mathbb{N}$ is just the real membership relation restricted to $\mathbb{N}$'s domain — i.e., $i_{\mathbb{N}}(\in) = \{\langle n_1, n_2\rangle \in N \times N \mid n_1 \in n_2\}$.

This terminology puts us in a position to understand the so-called "transitive submodel" version of Skolem's Paradox. Suppose that our favorite model of ZFC — i.e., $\mathbb{M}$ — is actually a countable *transitive* model.[35] Then there are four things we should immediately notice. First, transitivity takes care of all of the interpretation I/interpretation II type problems discussed in section 1. If $\mathbb{M}$ is transitive, then $\{x \mid x \in m\} = \{x \mid \mathbb{M} \models x \in m\}$ for every $m \in \mathbb{M}$. So, the fact that $\mathbb{M}$ is countable really does imply that each $m \in \mathbb{M}$ is also countable. Further, the equivalence of "$m \in \hat{m}$" with "$\mathbb{M} \models m \in \hat{m}$" implies that we don't need any $\in_{\hat{m}}$-style tricks to ensure that the syntax of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ line up appropriately.[36]

Second, the transitivity of $\mathbb{M}$ eliminates *one* of the semantic differences between $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ which we discussed in the last section. For any elements m_1 and m_2 in $\mathbb{M}$:

$$\mathbb{M} \models m_1 \in m_2 \quad \text{if and only if} \quad m_1 \in m_2.$$

As a result, any purely extensional differences between the way $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret the symbol "$\in$" vanish on transitive models. This means that we have to explain the transitive submodel version of Skolem's Paradox in terms of the way $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret their quantifiers.

Third, the transitivity of $\mathbb{M}$ ensures that $\mathbb{M}$ "gets it right" about a lot more than just the membership relation. Let me say that a relation R is *absolute* for transitive models if there is some formula $\Psi^R(\bar{x})$ such that for any transitive $\mathbb{N} \models \text{ZFC}$ and any $\bar{n} \in \mathbb{N}$:

$$R \text{ holds of } \bar{n} \iff \Psi^R_E(\bar{n}) \iff \Psi^R_{\mathbb{N}}(\bar{n}) \iff \mathbb{N} \models \Psi^R[\bar{n}].^{37}$$

---

[35] As I noted on page 620, the assumption that there exists a countable transitive model of ZFC is slightly stronger than the assumption that there exists an arbitrary model of ZFC. That being said, it's not a *particularly* strong assumption — it follows, for instance, from almost any standard large cardinal assumptions. Nonetheless, it is stronger anything we've assumed so far.

[36] Note that when $\mathbb{M}$ is transitive, $\hat{m} = \{x \mid \mathbb{M} \models x \in \hat{m}\}$. Hence, we can avoid writing things like $f : \omega \to \{x \mid \mathbb{M} \models x \in \hat{m}\}$ and just use the more perspicuous: $f : \omega \to \hat{m}$. I will use this later notation throughout the remainder of this section.

[37] Here, $\Psi^R_E(\bar{n})$ is just the "ordinary English" interpretation of $\Psi^R(\bar{n})$ and $\Psi^R_{\mathbb{N}}(\bar{n})$ is the in-

Clearly, the definition of transitivity ensures that the relation "is a member of" is absolute for transitive models. With a bit of work, we can show that the following are also absolute:

- f is a function; f is injective; f is surjective; f is bijective.
- $x = \text{Domain}(f)$; $x = \text{Range}(f)$.
- x is finite; x is infinite; x is an ordinal; x is a limit ordinal; $x = \omega$.

Hence, transitive models "know" quite a lot about the sets they contain. For a wide range of set-theoretic concepts, transitive models of ZFC pin these concepts down accurately (at least, that is, with respect to elements living in those models' domains).

Finally, the transitivity of $\mathbb{M}$ lets us determine just which symbol in $\Omega_{\mathbb{M}}(x)$ should "take the blame" for Skolem's Paradox. As we have already seen, the fact that $\mathbb{M}$ is transitive ensures that extensional differences between $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ must be located in the interpretation of "$\exists x$" (since differences involving the interpretation of "$\in$" have already been eliminated). Further, the above discussion of absoluteness provides us with the resources to isolate *just which instance* of "$\exists x$" really does the explanatory work.

To see this, we should first note that the class of concepts which are absolute for transitive models is rich enough to include the two-place relation "f is a bijection between ω and x." That is, there exists a formula $\Psi(f, x)$ such that for any transitive $\mathbb{N} \models \text{ZFC}$ and any $n_1, n_2 \in N$,

$$n_1 \text{ is a bijection between } \omega \text{ and } n_2 \iff \Psi_E(n_1, n_2) \iff \Psi_{\mathbb{N}}(n_1, n_2).^{38}$$

Further, the formula we've been calling $\Omega(x)$ is closely related to this formula $\Psi(f, x)$. In particular,

$$\Omega(x) \equiv_{df} \neg\exists f\, \Psi(f, x).$$

This gives us the technical machinery we need to explain where $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ really differ.

At the most general level, we can start with the fact that $\Omega_E(x)$ and $\Omega_M(x)$ clearly interpret the symbol "$\neg$" the same way: both make $\neg\phi$ true exactly when ϕ is false. Next, we note that the absoluteness of $\Psi(f, x)$ ensures that, for any particular $f, x \in \mathbb{M}$, the sentences $\Psi_E(f, x)$ and $\Psi_{\mathbb{M}}(f, x)$ are also extensionally

---

terpretation of this formula *at* $\mathbb{N}$. The notation is intended as a strict analog of the $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ notation introduce on page 626. I will use this type of notation freely throughout the remainder of this paper.

[38] To get this Ψ, note that we already have formulas $\Psi_1(f)$, $\Psi_2(x, f)$, $\Psi_3(y, f)$, and $\Psi_4(y)$, which capture, respectively, the concepts "f is a bijection," "$x = \text{Range}(f)$," $y = \text{Domain}(f)$, and "$y = \omega$." Note further that any transitive model of ZFC must already contain the real ω (since $\text{ZFC} \vdash \exists y\, y = \omega$ and the formula $y = \omega$ is absolute for transitive models). Therefore, the formula:

$$\Psi(f, x) \equiv \Psi_1(f) \wedge \Psi_2(x, f) \wedge \exists y\, (\Psi_3(y, f) \wedge \Psi_4(y))$$

accurately captures the concept "f is a bijection between ω and x."

equivalent. Hence, differences in the interpretation of symbols occurring *inside of* $\Psi(f,x)$ won't help to explain the failure of (†). When we combine these two facts, we see that the only significant difference between the semantics of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ involves the interpretation of the initial existential quantifier in $\Omega(x)$. For transitive models, therefore, the analysis of (†) given by the quantificational solution to Skolem's Paradox — i.e., the analysis which focuses solely on the range of the initial existential quantifier in $\Omega(x)$ — really does explain the failure of (†).

Let's take a closer look at this explanation by tracking it through a particular case. Since we already know that $\hat{m}$ provides a witness to the failure of (†) — i.e., that the conditional $\Omega_{\mathbb{M}}(\hat{m}) \Longrightarrow \Omega_E(\hat{m})$ is both false and an instantiation of (†) — we'll focus our attention there. Given what we already know about $\Psi(x,y)$, the following two facts are clear:

1. For any set f, $\Psi_E(f,\hat{m})$ is true if and only if f is a bijection between ω and $\hat{m}$.
2. For any $f \in \mathbb{M}$, $\Psi_{\mathbb{M}}(f,\hat{m})$ is true if and only if f is a bijection between ω and $\hat{m}$.

Further, the fact that $\mathbb{M}$ is countable entails that $\hat{m}$ is also countable. So, there really is a bijection $\hat{f} : \omega \to \hat{m}$.

Now, because $\hat{f}$ is a bijection between ω and $\hat{m}$, fact 1 entails that $\Psi_E(\hat{f},\hat{m})$ is true. So, since the quantifiers in $\Omega_E(\hat{m})$ range over the whole universe of sets — in particular, then, over a domain large enough to contain $\hat{f}$ — the expression $\exists\, f\, \Psi_E(f,\hat{m})$ must also be true. Hence, $\Omega_E(\hat{m}) \equiv \neg\exists\, f\, \Psi_E(f,\hat{m})$ must be false. In contrast, fact 2 only entails that $\mathbb{M}$ recognizes *those bijections which live in its domain.* That is, if some $f \in \mathbb{M}$ *is* a bijection between ω and $\hat{m}$, then $\mathbb{M}$ "knows" that it's a bijection between ω and $\hat{m}$, and if $f \in \mathbb{M}$ is *not* a bijection between ω and $\hat{m}$, then $\mathbb{M}$ "knows" that it's not. It so happens, however, that neither $\hat{f}$ nor any other bijection between ω and $\hat{m}$ lives in the domain of $\mathbb{M}$. Hence, the initial quantifier in $\Omega_{\mathbb{M}}(\hat{m})$ doesn't "see" any f for which $\Psi_{\mathbb{M}}(f,\hat{m})$ comes out true. As a result, $\exists\, f\, \Psi_{\mathbb{M}}(f,\hat{m})$ comes out false, and $\Omega_{\mathbb{M}}(\hat{m})$ comes out true.

This, then, gives us a more detailed explanation of the failure of (†). There are two things to note about this explanation. First, our ability to pin down the particular quantifier which accounts for the failure of (†) depends on the fact that $\mathbb{M}$ is a transitive model. It is *because* $\mathbb{M}$ is transitive that we know that the expressions $\Psi_E(f,x)$ and $\Psi_{\mathbb{M}}(f,x)$ are equivalent, and it is only because we know about this equivalence that we can isolate the initial "$\exists f$" in $\Omega(x)$ as the place where $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ really disagree. If $\mathbb{M}$ were *not* transitive, then we would have no reason for believing that $\Psi_E(f,x)$ and $\Psi_{\mathbb{M}}(f,x)$ are equivalent — in particular, we would have none of the absoluteness results from page 634. In *that* case, therefore, any of the instances of "$\in$" and "$\exists y$" which occur *inside of* $\Psi(f,x)$ could — at least in principle — explain the failure of (†) just as well as the initial "$\exists f$" in $\Omega(x)$ does.

Second, the clarity of this transitive model explanation helps, I think, to explain the popularity of the quantificational solution to Skolem's Paradox. As noted above, this is a case — and a very often cited case — where the quantificational solution *really does* explain what's going on in Skolem's Paradox. When we combine this with the fact — discussed on page 632 — that countable models *always do* exclude genuine bijections between ω and $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ from the range of their quantifiers, we can see the real virtues of the quantificational solution. Even if it doesn't provide a *complete* solution to Skolem's Paradox, it does provide an excellent partial solution — i.e., a solution which works perfectly well in some particular cases.

4 THE VICES OF QUANTIFICATION I

So, why doesn't the quantificational solution work in *all* cases? Why does it fail as a *general* solution to Skolem's Paradox? To answer these questions, recall one of the roles that transitivity played in the last section. By making $\mathbb{M}$ transitive, we ensured that if $\mathbb{M}$ *contains* a bijection $f : \omega \to \hat{m}$, then $\mathbb{M}$ also *recognizes* f as a bijection from ω to $\hat{m}$. This was the point of our discussion of absoluteness, and it played a key role in allowing us to isolate a particular quantifier as the one which "explained" Skolem's Paradox. To generate an example where this kind of analysis breaks down, therefore, we should start by looking for a model which contains various bijections without recognizing them as bijections.

Fortunately, it's relatively easy to find such a model. The idea is to start with a transitive model, $\mathbb{N}$, and then use our footnote 9 trick to replace some element of $\mathbb{N}$ with a bijection of the relevant sort (while leaving enough other things fixed that our new model, $\mathbb{M}$, doesn't recognize this new element *as* a bijection of the relevant sort). More formally, let $\mathbb{N}$ be a countable, transitive model of ZFC, let $\hat{m}$ be an element of $\mathbb{N}$ such that $\mathbb{N} \models \Omega[\hat{m}]$, and let $\hat{n}$ be an element of $\mathbb{N}$ such that $\text{Rank}(\hat{n}) > \text{Rank}(\hat{m}) + \omega$.[39] Now, since $\mathbb{N}$ is countable and transitive, the set $\hat{m} = \{x \mid \mathbb{N} \models x \in \hat{m}\}$ is also countable; so, there exists a bijection $\hat{f} : \omega \to \hat{m}$. Next, we define a function $\sigma : N \to (N \setminus \{\hat{n}\}) \cup \{\hat{f}\}$ such that:

$$\sigma(n) = \begin{cases} n & \text{if } n \neq \hat{n} \\ \hat{f} & \text{if } n = \hat{n} \end{cases}$$

This function σ allows us — by means of the trick described in footnote 9 — to construct a new model, $\mathbb{M}$, such that 1.) $\text{Domain}(\mathbb{M}) = (N \setminus \{\hat{n}\}) \cup \{\hat{f}\}$ and 2.) σ is an isomorphism between $\mathbb{N}$ and $\mathbb{M}$.

From a technical perspective, this new model has four nice properties. First, because σ is an isomorphism between $\mathbb{N}$ and $\mathbb{M}$, $\mathbb{M}$ satisfies the same sentences as $\mathbb{N}$ did; in particular, therefore, $\mathbb{M} \models \mathrm{ZFC}$. Second, because $\mathbb{N} \models \Omega[\hat{m}]$ and $\sigma : \mathbb{N} \to \mathbb{M}$ is an isomorphism such that $\sigma(\hat{m}) = \hat{m}$, $\mathbb{M} \models \Omega[\hat{m}]$ as well. Third, because we chose $\hat{n}$ from a different "part" of $\mathbb{N}$ than $\hat{m}$, the equivalence $\hat{m} = \{x \mid x \in \hat{m}\} = \{x \mid \mathbb{M} \models x \in \hat{m}\}$ caries over from the transitive model case (see fn. 36). Finally, because $\hat{f} \in \mathbb{M}$, $\mathbb{M}$ *contains* a function which witnesses the fact that $\hat{m}$ is countable. Although $\mathbb{M}$ does not *recognize* this function in the right sort of way — as indicated by the fact that $\mathbb{M} \models \Omega[\hat{m}]$ — $\mathbb{M}$ *does contain* the relevant function.[40]

---

[39] A remark on this choice of $\hat{n}$ is in order. Basically, I have chosen $\hat{n}$ so as to ensure that it does not live in the same "part" of $\mathbb{N}$ as $\hat{m}$ does. In particular, $\hat{n}$ is not a member of either $\hat{m}$ or ω. What's more, $\hat{n}$ isn't equal to any ordered pair of the form $\langle n_1, n_2 \rangle$, where $n_1 \in \hat{m}$ and $n_2 \in \omega$, nor is $\hat{n}$ equal to a collection of such ordered pairs. As a result, we can manipulate $\hat{n}$ in various ways without modifying the parts of $\mathbb{N}$ which directly involve $\hat{m}$, ω, and $\omega \times \hat{m}$. The significance of this choice of $\hat{n}$ will become clear as my argument progresses.

From a more philosophical perspective, this example brings out two things about Skolem's Paradox. First, it provides an example where the quantificational analysis of Skolem's Paradox starts to break down. Informally, it's no longer true to say that the quantifiers in $\Omega_{\mathbb{M}}(\hat{m})$ don't "know" about any bijections between ω and $\{x \mid \mathbb{M} \models x \in \hat{m}\}$ (while those in $\Omega_E(\hat{m})$ do know about such bijections). More formally, the key absoluteness result on which the analysis of the last section depended — i.e., that for $\Psi(f, x)$ — doesn't hold in the current context. In this context, $\Psi_E(\hat{f}, \hat{m})$ is true, but $\Psi_{\mathbb{M}}(\hat{f}, \hat{x})$ is not.

Second — and more strongly — I think this case provides an example where Skolem's Paradox can't be explained by quantifier-ranges at all. To see this, we can begin by taking a closer look at the non-absoluteness of $\Psi(f, x)$. On the one hand, because $\mathbb{M} \models \Omega[\hat{m}]$, we know that $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$ must be false. On the other hand, the fact that $\hat{f}$ really is a bijection between ω and $\hat{m}$ entails that $\Psi_E(\hat{f}, \hat{m})$ must be true. These facts, together with the fact that both $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ "know" about $\hat{f}$ — i.e., the fact that $\hat{f}$ is within the range of the quantifiers of both $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ — suggest that any differences between $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ lie *inside of* $\Psi(f, \hat{m})$, rather than in the way $\Omega_{\mathbb{M}}(x)$ and $\Omega_E(x)$ interpret their initial existential quantifiers.

Following this line of thought, let's look even more closely at $\Psi(\hat{f}, \hat{m})$. Abbreviating wildly, we get:

$$\begin{aligned} \Psi(\hat{f}, \hat{m}) \quad \equiv_{df} \quad & \forall x \in \hat{f}\, [x \in \omega \times \hat{m}] \\ & \wedge\, \forall x \in \omega\ \exists! y \in \hat{m}\, [\langle x, y \rangle \in \hat{f}] \\ & \wedge\, \forall y \in \hat{m}\ \exists! x \in \omega\, [\langle x, y \rangle \in \hat{f}]. \end{aligned}$$

Now, if we examine this formula closely, we will find that many of its subformulas receive equivalent interpretations under the semantics of $\Psi_E(\hat{f}, \hat{m})$ and those of $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$; to use our earlier jargon, these subformulas are *absolute* between V and $\mathbb{M}$. In particular, we should observe that for any set s:[41]

- $s \in \hat{m} \iff \mathbb{M} \models s \in \hat{m}$.

---

[40] Of course, there are other bijections $f : \omega \to \hat{m}$ which don't live in the domain of $\mathbb{M}$ (since, as before, there are $2^{\aleph_0}$ different bijections between ω and $\hat{m}$, and not all of them can live in the countable domain of $\mathbb{M}$). I will discuss these other bijections in section 5.

[41] Note, here, that it's our original choice of $\hat{n}$ which ensures that the following claims are true. By choosing $\hat{n}$ to be in a different "part" of $\mathbb{N}$ than $\hat{m}$ and ω were, we ensured that replacing $\hat{n}$ with $\hat{f}$ does not effect the following properties. See footnote 39 for more on this.

- $s \in \omega \Longleftrightarrow \mathbb{M} \models s \in \omega$.
- $s \in \omega \times \hat{m} \Longleftrightarrow \mathbb{M} \models s \in \omega \times \hat{m}$.
- If $s_1 \in \omega$ and $s_2 \in \hat{m}$, then $s = \langle s_1, s_2 \rangle \Longleftrightarrow \mathbb{M} \models s = \langle s_1, s_2 \rangle$.

Finally, we should observe that, with respect to the sets that are actually relevant to the truth or falsity of $\Psi(\hat{f}, \hat{m})$, there's a fair bit of agreement between the quantifiers in $\Psi_E(\hat{f}, \hat{m})$ and those in $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$. In particular, every $s \in \hat{n} \cup \hat{m} \cup \omega \cup (\omega \times \hat{m})$ lives in the range of both sets of quantifiers.

Keeping these observations in mind, we can distinguish three kinds of differences between the semantics of $\Psi_E(\hat{f}, \hat{m})$ and those of $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$. **First**, there are differences that occur *within* subformulas that are, themselves, absolute between V and $\mathbb{M}$ — e.g., formal differences in the interpretation of quantifiers within expressions like "$x \in \omega \times \hat{m}$." **Second**, there are differences in the interpretation of quantifiers where 1.) these quantifiers are explicitly bounded as they occur in $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$ and 2.) these quantifiers have ranges, both as they occur in $\Psi_E(\hat{f}, \hat{m})$ and as they occur in $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$, which include every element in *either* of the relevant bounding sets. So, for instance, the initial quantifier in "$\forall x \in \hat{f}\, [x \in \omega \times \hat{m}]$" is bounded by the expression "$\in \hat{f}$"; but the ranges of the quantifiers in both $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$ are large enough to include $\{x \mid x \in \hat{f}\} \cup \{y \mid \mathbb{M} \models y \in \hat{f}\}$. Similarly, the initial quantifiers in "$\forall x \in \hat{m}\ \exists! y \in \omega\, [\langle x, y \rangle \in \hat{f}\,]$" are bounded by "$\in \hat{m}$" and "$\in \omega$" respectively; but the quantifiers in both $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$ range over $\hat{m} \cup \{y \mid \mathbb{M} \models y \in \hat{m}\}$ and $\omega \cup \{y \mid \mathbb{M} \models y \in \omega\}$. **Third**, there are differences in the interpretation of the membership sign in the three instances of the expression "$\in \hat{f}$" which occur in $\Psi(\hat{f}, \hat{m})$ — i.e., one instance of "$x \in \hat{f}$" and two of "$\langle x, y \rangle \in \hat{f}$."

Clearly, neither of the first two kinds of difference can explain the difference in truth-value between $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$. Because differences of the first kind are isolated within subformulas whose truth-values are constant between $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$, these differences cannot be where $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$ ultimately diverge. Similarly for differences of the second kind. Although the quantifiers in $\Psi_E(\hat{f}, \hat{m})$ range over a *larger* domain than those in $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$, none of the "extra" objects within the range of $\Psi_E(\hat{f}, \hat{m})$'s quantifiers are relevant to the truth-values of formulas like $\forall x \in \hat{f}\, [x \in \omega \times \hat{m}]$ or $\forall x \in \hat{m}\ \exists! y \in \omega\, [\langle x, y \rangle \in \hat{f}]$ (whether these formulas are interpreted after the fashion of Ψ_E or of $\Psi_{\mathbb{M}}$). Hence, none of these differences in quantifier-ranges can explain the final difference in truth-value between $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$.

This, therefore, brings us back to the third difference between $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$ — their differing interpretations of the membership sign in the expression "$\in \hat{f}$." As this is the only difference which is not covered by cases 1 and 2, it must be the one which explains the difference in truth-values between $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$. Further, this explanation is relatively intuitive. The notion captured by $\Psi_E(x, y)$ — that "x is a bijection between ω and y" — is a notion that's *defined in terms of* the members of x. So, since $\mathbb{M}$ doesn't know about the *real* members of $\hat{f}$ (recall, the things $\mathbb{M}$ thinks are members of $\hat{f}$ are really members of $\hat{n}$), $\mathbb{M}$ doesn't know that $\hat{f}$ is a bijection between ω and $\hat{m}$. Hence, it's not surprising that $\mathbb{M}$ fails to satisfy the formula $\Psi(\hat{f}, \hat{m})$. It's a simple consequence of the discrepancy between $\mathbb{M}$'s understanding of the membership relation on $\hat{f}$ and the real membership relation on $\hat{f}$.

This gives us an analysis of the failure of (†) which is quite different from the one given in section 3. There, the explanation for the failure of (†) involved the differing interpretations which $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$ give to their quantifiers (and, in particular, to one specific quantifier). Here, the explanation depends on the way these expressions interpret the symbol "$\in$" in the embedded formula $\Psi(f, \hat{m})$ (and, again, we can limit our attention to a few specific instances of "$\in$"). Of course, there are other places where the semantics $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$ differ formally (that's true in the transitive submodel case as well), but these are the differences which really *explain* the present version of Skolem's Paradox.

This example gives an initial indication as to why the standard, quantificational solution to Skolem's Paradox is inadequate. In particular, it shows that an analysis of Skolem's Paradox which works fine for transitive models *does not* work for all models (e.g., because we may lose the absoluteness of $\Psi(f, x)$). Further, I've made a preliminary argument for the claim that this new example can't be explained by looking at quantifier-ranges at all. In the next section, I'll bolster this argument by looking at two more examples. Along the way, I'll try to dispel a worry that the argument of this section may have occasioned. I'll end with some general remarks about Skolem's Paradox.

5 THE VICES OF QUANTIFICATION II

Let's start with a possible concern about the analysis of the last section. In giving the analysis, I noted that the bijection $\hat{f}$ lived in the domain of the quantifiers of both $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$, and I then focused my attention on the difference in truth value between $\Psi_E(\hat{f}, \hat{m})$ and $\Psi_{\mathbb{M}}(\hat{f}, \hat{m})$. There are, however, $2^{\aleph_0}$ other bijections $g : \omega \to \hat{m}$ which don't live in the domain of $\mathbb{M}$. Why can't one of these other bijections "explain" the fact that $\Omega_E(\hat{m})$ is false while $\Omega_{\mathbb{M}}(\hat{m})$ is true? Why, in short, can't the difference between $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$ still be explained — if only partially — by the way these sentences interpret their initial existential quantifiers?

There are, I think, three ways of responding to these questions. First, we should note that there is no *particular* $g : \omega \to \hat{m}$ which plays a special role in explaining the differences between $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$. After all, for any particular g, we can easily build a new version of $\mathbb{M}$ which contains that g — i.e., by substituting it for $\hat{f}$ in the preceding construction. Indeed, a trivial modification of that construction allows us to include countably many bijections $g : \omega \to \hat{m}$ within the domain of

$\mathbb{M}$.[42] So, there's no sense in which we've somehow used the *wrong* $\hat{f}$ in building our model $\mathbb{M}$.

Second, it's hard to see how some $g : \omega \to \hat{m}$ *could* explain the difference between $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$. Presumably, the thought goes something like this. In the case of $\Omega_E(\hat{m})$, the quantifiers range over a domain large enough to include the relevant g's. Further, the semantics of $\Omega_E(\hat{m})$ recognize these g's as bijections from ω to $\hat{m}$. So, $\exists f\, \Psi_E(f, \hat{m})$ comes out true, and $\Omega_E(\hat{m})$ comes out false. The idea, then, is that this kind of analysis should *almost* work for $\Omega_{\mathbb{M}}(\hat{m})$. *If* $\mathbb{M}$ knew about some bijection $g : \omega \to \hat{m}$, *then* $\mathbb{M}$ would recognize g as a bijection from ω to $\hat{m}$. As a result, $\mathbb{M}$ would satisfy some formula of the form $\Psi(g, \hat{m})$; so, it would also satisfy $\exists f\, \Psi(f, \hat{m})$, and it would fail to satisfy $\Omega(\hat{m})$. In short, if the quantifiers in $\Omega_{\mathbb{M}}(\hat{m})$ only knew about the same functions that the quantifiers in $\Omega_E(\hat{m})$ know about, then the analysis of $\Omega_{\mathbb{M}}(\hat{m})$ would run exactly parallel to that of $\Omega_E(\hat{m})$.

Why, though, should we believe any of this? After all, $\mathbb{M}$ *already does* contain one bijection $\hat{f} : \omega \to \hat{m}$, and $\mathbb{M}$ *doesn't* recognize $\hat{f}$ as a bijection (or, at least, not as a bijection between ω and $\hat{m}$). Why should we think $\mathbb{M}$ would do any *better* when it comes to other bijections? In the transitive model case, our absoluteness results ensured that $\mathbb{M}$ got bijections right — that if $\mathbb{M}$ knew about some bijection $g : m \to n$, then $\mathbb{M}$ recognized g as a bijection from m to n. So, it was at least superficially plausible to think that if $\mathbb{M}$ *could know* about some new $g : \omega \to \hat{m}$, then $\mathbb{M}$ would properly recognize g as a bijection from ω to $\hat{m}$.[43] But, once $\mathbb{M}$ misidentifies *one* bijection — i.e., $\hat{f}$ — then there's no particular reason to think it should do better with respect to *other* bijections.

These first two points suggest that the questions raised at the beginning of this section are not well-motivated. Unlike in the transitive model case, there may simply be no coherent story about how initial quantifiers *could* help to explain the version of Skolem's Paradox we're now considering. If so, then our solution to this version of Skolem's Paradox is going to have to look quite a bit different from the quantifier-oriented solution presented in section 3. And, while this doesn't *directly* show that the membership-oriented solution presented in the last section is correct, it does lend that solution a good deal of *indirect* support.[44]

This brings me to a third point. As we saw earlier, part of the appeal of the quantificational solution to Skolem's Paradox comes from the fact that, for any countable $\mathbb{M}$ and any $\hat{m} \in \mathbb{M}$, there are $2^{\aleph_0}$ bijections $g : \omega \to \{m \mid \mathbb{M} \models m \in \hat{m}\}$

---

[42] I.e., we just use σ to replace countably many elements of $\mathbb{N}$ (all with sufficiently high rank) with new bijections $g : \omega \to \hat{m}$.

[43] Of course, it's hard to know how to evaluate this kind of subjunctive conditional, since its antecedent is *necessarily* false. We can, however, say the following: *if* $\mathbb{M}$ were extended to a larger transitive model, $\mathbb{M}'$, such that $\mathbb{M}'$ contained a bijection $g : \omega \to \hat{m}$, *then* $\mathbb{M}'$ would recognize this g as a bijection between ω and $\hat{m}$.

[44] Recall, here, that the discussion of absoluteness and bounded quantification on pp 637–638 showed that most of the symbols in $\Omega(x)$ are irrelevant to explaining this version of Skolem's Paradox. If, as suggested above, the initial quantifier in $\Omega(x)$ is also irrelevant, then the instances of "$\in$" isolated on page 638 provide the only possible focus for our explanation.

which don't live in the domain of $\mathbb{M}$. Clearly, there's no way of formulating a version of Skolem's Paradox which allows us to evade this fact.[45] We can, however, formulate puzzles that are closely *analogous* to Skolem's Paradox and which *do* allow us to evade this fact. Further, the solution to these later puzzles follows precisely the lines given in the last section. This leads me, once again, to think that the solution from the last section is correct (and that the fact about "missing" bijections is largely a red herring in our present context).

Let's look at two of these analogous puzzles. The first involves the comparison of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ where $\mathbb{M}$ is a, suitably chosen, *uncountable* model of ZFC. As usual, we'll start by letting $\mathbb{N}$ be a countable, transitive model of ZFC. Applying a theorem of Keisler and Morley, we generate a model $\mathbb{N}'$ such that 1.) $\mathbb{N}'$ is an elementary end extension of $\mathbb{N}$ and 2.) $|N'| = 2^{\aleph_0}$.[46] Now, let $\hat{m}$ be an element of $\mathbb{N}$ such that $\mathbb{N} \models \Omega[\hat{m}]$, and let $\mathbb{X} = \{g : \omega \to \hat{m} \mid g \text{ is a bijection}\}$ (so, $\mathbb{X}$ is the set of *real* bijections between ω and $\hat{m}$). Finally, using the fact that $|N'| = 2^{\aleph_0}$, we build a bijection $\sigma : N' \to N' \cup \mathbb{X}$ such that $\sigma \restriction N = \mathrm{Id}$; we let $\mathbb{M}$ be the model induced by this σ — i.e., induced in the manner described in footnote 9.

At the end of this construction, our new model, $\mathbb{M}$, has five nice properties: 1.) $\mathbb{M} \models \text{ZFC}$, 2.) $\mathbb{M} \models \Omega[\hat{m}]$, 3.) $\hat{m} = \{x \mid \mathbb{M} \models x \in \hat{m}\}$, 4.) $\hat{m}$ is countable, and 5.) every real bijection, $g : \omega \to \hat{m}$, is actually a member of $\mathbb{M}$. Here, 1 follows from the fact that $\mathbb{N}$ satisfies ZFC, together with the fact that $\mathbb{N}'$ is an elementary extension of $\mathbb{N}$ and that $\sigma : \mathbb{N}' \to \mathbb{M}$ is an isomorphism. 2 follows from the same facts, along with the fact that $\sigma(\hat{m}) = \hat{m}$. 3 and 4 follow from the fact that $\mathbb{N}$ is countable and transitive, together with the fact that $\mathbb{N}'$ is an end extension of $\mathbb{N}$ and that $\sigma \restriction N = \mathrm{Id}$. Finally, 5 follows from our choice of $\mathbb{X}$, together with the fact that $\mathbb{X} \subset M$.

There are two things to notice about all this machinery. First, properties 1–4 give rise to an obvious analog of Skolem's Paradox. After all, $\mathbb{M}$ is a model of ZFC which satisfies $\Omega[\hat{m}]$ (properties 1 and 2), despite the fact that the set $\hat{m} = \{x \mid \mathbb{M} \models x \in \hat{m}\}$ is only countable (properties 3 and 4). So, just as in our previous examples, $\Omega_{\mathbb{M}}(\hat{m})$ is true, and $\Omega_E(\hat{m})$ is false. Second, this analog of Skolem's Paradox neatly evades the concerns raised at the beginning of this section. After all, fact 5 ensures that *all* of the $2^{\aleph_0}$ bijections $g : \omega \to \hat{m}$ which witness the falsity of $\Omega_E(\hat{m})$ live within the domain of $\mathbb{M}$.

Indeed, we can go a bit further than this. It's pretty clear that any solution to this analog of Skolem's Paradox *has* to run parallel to the solution sketched on pp

---

[45] As a matter of classification, I take it that "Skolem's Paradox" always involves the comparison of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$, where $\mathbb{M}$ is a *countable* model. Hence, the fact mentioned in the main text will always be present (though, as I have argued, it may not always be *relevant*).

[46] Some clarification is probably in order here. To say that $\mathbb{N}'$ is an elementary extension of $\mathbb{N}$ means that for any formula $\phi(\bar{x})$ and any sequence $\bar{n} \in \mathbb{N}$, $\mathbb{N}' \models \phi[\bar{n}] \iff \mathbb{N} \models \phi[\bar{n}]$. In particular, then, the fact that $\mathbb{N}$ satisfies ZFC entails that $\mathbb{N}'$ also satisfies ZFC; further, for any $n \in \mathbb{N}$, $\mathbb{N}' \models \Omega[n] \iff \mathbb{N} \models \Omega[n]$. To say that $\mathbb{N}'$ is an end extension of $\mathbb{N}$ means that for every $n \in \mathbb{N}$, $\{x \mid \mathbb{N} \models x \in n\} = \{x \mid \mathbb{N}' \models x \in n\}$; so, moving to $\mathbb{N}'$ doesn't involve adding "new" elements to old members of $\mathbb{N}$. Given this, Keisler and Morley proved (see [Keisler and Morley, 1968]) that any countable model of ZFC has elementary end extensions of arbitrarily large cardinality.

637–638. Since *every* bijection which witnesses the falsity of $\Omega_E(\hat{m})$ — i.e., every g which makes $\Psi_E(g, \hat{m})$ come out true — is contained within the domain of $\mathbb{M}$, the difference in truth-value between $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$ can't be explained by looking at the interpretation of the initial quantifiers in $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$. Nor can we explain $\mathbb{M}$'s failure to "recognize" elements of $\mathbb{X}$ *as* bijections between ω and $\hat{m}$ by appealing to way $\mathbb{M}$ interprets the quantifiers in the embedded formula $\Psi(x, \hat{m})$. To be sure, the quantifiers in formulas like $\Psi_E(g, \hat{m})$ *do* range over a larger domain than those in $\Psi_{\mathbb{M}}(g, \hat{m})$. As before, however, every set which is relevant to the truth of $\Psi_E(g, \hat{m})$ is a member of $\omega \cup \hat{m} \cup (\omega \times \hat{m})$. These sets *are* within the range of the quantifiers in $\Psi_{\mathbb{M}}(g, \hat{m})$, and *every* set in the range of the quantifiers of $\Psi_{\mathbb{M}}(g, \hat{m})$ is within the range of the quantifiers in $\Psi_E(g, \hat{m})$.[47] Hence, the only remaining explanation for the difference in truth-values between $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$ stems from the way these formulas interpret the symbol "$\in$."[48]

This, therefore, gives us an analog of Skolem's Paradox whose solution has to follow the lines sketched in the last section. I'll close this section with a second such analog. Whereas my first example involved the comparison of $\Omega_E(x)$ and $\Omega_{\mathbb{M}}(x)$ for an uncountable model, $\mathbb{M}$, this one will compare $\Omega_{\mathbb{N}}(x)$ and $\Omega_{\mathbb{M}}(x)$ where $\mathbb{N}$ and $\mathbb{M}$ are both countable. As before, we can start by letting $\mathbb{N}$ be a countable, transitive model of ZFC. Next, let $\mathbb{N}' = \mathbb{N}[G]$ be a generic extension of $\mathbb{N}$ such that $\omega_1^{\mathbb{N}}$ has been "collapsed" to have cardinality $\aleph_0$.[49] Given this, let $X = \{n \in \mathbb{N} \mid \mathbb{N} \models$ "Rank$(n) < \omega_\omega$"$\}$, and let $\sigma : N' \to N$ be a bijection such that $\sigma \restriction X = \mathrm{Id}$. Finally, using our trick from footnote 9, we can define a new membership relation on the domain of $\mathbb{N}$ such that σ becomes an isomorphism between $\mathbb{N}'$ and $\mathbb{M}$ (where $\mathbb{M}$ is the new model canonically induced by this bijection).

At this point, we are in a position to formulate a puzzle very much like Skolem's Paradox except that it holds *between* $\mathbb{N}$ and $\mathbb{M}$ (rather than between $\mathbb{N}$ and the set-theoretic universe). To begin, note that the fact that $\sigma \restriction X = \mathrm{Id}$ ensures that $\mathbb{N}$ and $\mathbb{M}$ agree about the membership relation on $\omega_1^{\mathbb{N}}$. That is,

$$\{x \mid \mathbb{N} \models x \in \omega_1^{\mathbb{N}}\} = \{x \mid \mathbb{N}' \models x \in \omega_1^{\mathbb{N}}\} = \{x \mid \mathbb{M} \models x \in \omega_1^{\mathbb{N}}\}.^{50}$$

However, $\mathbb{N}$ and $\mathbb{M}$ do not agree about the *countability* of $\omega_1^{\mathbb{N}}$. On the one hand, $\mathbb{N} \models \Omega[\omega_1^{\mathbb{N}}]$. On the other hand, our forcing construction ensures that $\mathbb{N}' \models \neg\Omega[\omega_1^{\mathbb{N}}]$

---

[47] As in our original example, this point can be put in terms of "bounding sets." For any particular $g : \omega \to \hat{m}$, all of the quantifiers in $\Psi(g, \hat{m})$ are bounded by sets like g, $\hat{m}$, ω, and $\omega \times \hat{m}$. Since the elements of these bounding sets all live in the domain of $\mathbb{M}$, the quantifiers in $\Psi_{\mathbb{M}}(f, \hat{m})$ will "know" about all these elements.

[48] Again, as in the original example, a careful analysis allows us to isolate three specific instances of "$\in$" which do all the explanatory work. Since there's nothing new going on in this particular case, I omit the details of this further analysis.

[49] The details of this construction are too complicated to explain fully here. The relevant facts about $\mathbb{N}'$ are these: 1.) $\mathbb{N}'$ is a countable, transitive model of ZFC, 2.) $\mathbb{N}'$ is an "end extension" of $\mathbb{N}$ (cf. footnote 46), and 3.) $\mathbb{N}' \models$ "$\omega_1^{\mathbb{N}}$ is countable". Note that 3 entails that $\mathbb{N}' \models \neg\Omega[\omega_1^{\mathbb{N}}]$ and that 1–3 together entail that $\mathbb{N}'$ contains some (real) bijection $g : \omega \to \omega_1^{\mathbb{N}}$. Further details about this construction can be found in chapter 7 of [Kunnen, 1980] or chapter 3 of [Jech, 1978].

[50] The first of these equalities follows from the fact that $\mathbb{N}'$ is an end extension of $\mathbb{N}$. The second follows from the fact $\sigma \restriction X = \mathrm{Id}$. Basically, the fact that σ is the identity in the "neighborhood" of $\omega_1^{\mathbb{N}}$ ensures that "locally-definable" properties of $\omega_1^{\mathbb{N}}$ will be absolute between $\mathbb{N}$ and $\mathbb{M}$ (and,

(see footnote 49). So, the fact that $\sigma : \mathbb{N}' \to \mathbb{M}$ is an isomorphism, together with the fact that $\sigma(\omega_1^{\mathbb{N}}) = \omega_1^{\mathbb{N}}$, ensures that $\mathbb{M} \models \neg\Omega[\omega_1^{\mathbb{N}}]$ as well.

This gives us a simple analog of Skolem's Paradox: even though $\mathbb{N}$ and $\mathbb{M}$ agree about the members of $\omega_1^{\mathbb{N}}$, the expression $\Omega_{\mathbb{N}}(\omega_1^{\mathbb{N}})$ comes out true, and the expression $\Omega_{\mathbb{M}}(\omega_1^{\mathbb{N}})$ comes out false. Further, there's no possibility of explaining this discrepancy by appealing to the differing ways $\Omega_{\mathbb{N}}(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret their quantifiers. Since $\mathbb{N}$ and $\mathbb{M}$ have the same domain, $\Omega_{\mathbb{N}}(x)$ and $\Omega_{\mathbb{M}}(x)$ interpret their quantifiers *in exactly the same way.* Hence, any difference in truth-value between $\Omega_{\mathbb{M}}(\omega_1^{\mathbb{N}})$ and $\Omega_{\mathbb{N}'}(\omega_1^{\mathbb{N}})$ *must* be explained in terms of the differing ways $\mathbb{N}$ and $\mathbb{M}$ interpret the symbol "$\in$." In short, this is a case where the *only possible* solution to our puzzle follows the membership-oriented lines sketched in section 4.[51]

This, then, explains why I think the concerns raised at the beginning of this section are misguided. Although it's certainly true that, given any countable $\mathbb{M} \models \text{ZFC}$ and any $\hat{m} \in \mathbb{M}$, there will be $2^{\aleph_0}$ bijections $g : \omega \to \{x \mid \mathbb{M} \models x \in \hat{m}\}$ which don't live in the domain of $\mathbb{M}$, it's not at all clear that these bijections are relevant to the solution of (all versions of) Skolem's Paradox. For the version of Skolem's Paradox discussed in the last section, it's unclear *how* these bijections are supposed to explain the difference in truth-value between $\Omega_E(\hat{m})$ and $\Omega_{\mathbb{M}}(\hat{m})$ (while it's quite clear how certain instances of "$\in$" could do this explanatory work). Further, there are cases which involve the *same kind* of phenomena as Skolem's Paradox where the corresponding bijections simply don't exist. Given this, we should be cautious about insisting that Skolem's Paradox has a single, uniform explanation which can be formulated in terms of quantifier ranges. Although such quantificational solutions work well in certain cases — e.g., the transitive model case — there are other solutions which work better when we turn to more complicated cases — e.g., the case discussed in section 4.

6 A FEW CONCLUDING REMARKS

In the first five sections of this paper, I provided a tour through (some of) the mathematical issues involved in Skolem's Paradox. I looked at what it takes to make this paradox "look plausible," what we need to "solve" the paradox, what *different* solutions are appropriate for different versions of the paradox, etc. In this section, I want to step back and take a somewhat broader view. I'll start by highlighting a few of the main points — both technical and philosophical — from

---

for that matter, between either of these models and V). With a little work, we can show that *all* of the properties listed on pp. 637–638 are absolute between $\mathbb{N}$ and $\mathbb{M}$, when we put $\omega_1^{\mathbb{N}}$ in place of our previous $\hat{m}$. In particular, then, "$x = \omega$" and "$x \in \omega \times \omega_1^{\mathbb{N}}$" will both be absolute.

[51] In fact, the solution of this puzzle follows our previous solution rather closely. By letting $\mathbb{N}'$ be an end extension of $\mathbb{N}$ and then setting $\sigma \restriction X = \text{Id}$, we ensured that the vast majority of the symbols in $\Omega(\omega_1^{\mathbb{N}})$ occur within subformulas that are absolute between $\mathbb{N}$ and $\mathbb{M}$. This fact, together with the fact that $\mathbb{N}$ and $\mathbb{M}$ agree on the interpretation of their quantifiers, entails that any differences between $\Omega_{\mathbb{N}}(\omega_1^{\mathbb{N}})$ and $\Omega_{\mathbb{M}}(\omega_1^{\mathbb{N}})$ must be explained by ways these formulas interpret the usual three instances of "$\in$" — i.e., the three instances highlighted in the "$\in f$" clauses in the formulation of $\Psi(f, x)$ on page 637.

the preceding sections. I'll then examine what I take to be the main philosophical worry which the preceding discussion may have provoked.

From a technical perspective, there are two points I want to emphasize. First, it's a lot harder to formulate a plausible-looking version of Skolem's Paradox than it may seem to be at first. To make Skolem's Paradox look plausible, we need to exercise some care in choosing the countable model in terms of which the paradox is formulated (in general, an arbitrary countable model of ZFC won't do the job). We also need to think about just *how* we explicate our ordinary English notion of countability. Finally, we need to ensure that the choices we make with regard to these first two issues *fit together* appropriately: if, for instance, we use a model with a designated $\in_{\hat{m}}$-relation to formulate the paradox, then we need to explicate our notion of countability in terms of that relation. So, even getting a superficially plausible version of the paradox onto the table may require some careful technical work.

Second, Skolem's Paradox isn't *just* a puzzle concerning the interpretation of quantifiers. To be sure, there are some versions of the paradox which are best solved by looking at the way first-order models interpret quantification — e.g., the transitive submodel argument discussed in section 3. But, there are other versions of the paradox which require quite different solutions — e.g., the versions examined in sections 4–5. Given this, we should resist the idea that Skolem's Paradox has a completely general explanation which can be formulated in terms of quantifier ranges. Indeed, if we're really looking for a *general* solution to Skolem's Paradox — a solution which applies to *all* formulations of that paradox — then I doubt we can find one which is much more specific than the "generic solution" of section 2.

Let me make a comment about this second point. For expository reasons, sections 3–5 focused on cases where we could pretty easily isolate the specific symbols whose interpretation served to "explain" Skolem's Paradox (instances of "$\exists x$" in 3, and instances of "$\in$" in 4–5). I don't, however, think that all cases are quite this simple. As we saw on page 617, $\Omega(x)$ is an incredibly long formula, so there are many different symbols which can, in the context of specific models, "take the blame" for particular instances of Skolem's Paradox. Further, there are cases where Skolem's Paradox can't be pinned on specific symbols at all — i.e., cases where the paradox turns on a complicated interplay between the interpretation of several different symbols.[52] So, unless we're willing to accept a wildly disjunctive solution — potentially one with $2^{\aleph_0}$ cases — I don't think we can get a general solution to Skolem's Paradox which is more specific than that given in section 2.[53]

---

[52] To explore this possibility, the reader is encouraged to think through the case where κ is an inaccessible cardinal and $\mathbb{M}$ is a countable elementary submodel of $\langle V_\kappa, \in \rangle$. Let $\hat{m} = \aleph_{17}$, and analyze the version of Skolem's Paradox which results from the fact that $\mathbb{M} \models \Omega[\hat{m}]$.

[53] I do, however, think there's a lot to learn from tracking Skolem's Paradox through the details of various specific models. In doing so, we learn about the particular pathologies which give rise to things like Skolem's Paradox, about the strengths and weaknesses of first-order formulations of set theory, and about the fine details of our (various) conceptions of cardinality. There are a lot of open questions here, and I encourage philosophers — or, at least, those philosophers who are as fascinated by Skolem's Paradox as I am — to spend more time exploring this paradox in

So much, then, for the technical issues. On the more philosophical side, there are also two points worth emphasizing. First, there shouldn't be anything too surprising about Skolem's Paradox. At a general level, we can isolate clear differences between the semantics of ordinary English set theory and the model-theoretic semantics of formulas like $\Omega(x)$ — e.g., in their interpretation of symbols like "$\in$" and "$\exists$" — and we can see how Skolem's Paradox turns on an equivocation between these two kinds of semantics. At a more local level, when we track these differences through the details of particular models, we can often isolate just which symbols really give rise to Skolem's Paradox, and we can explain *how* the interpretation of these symbols gives rise to the paradox. Given all this, the paradox itself should no longer seem very puzzling.

Second, the argument of sections 4–5 should lead us to be cautious about any philosophical analysis of Skolem's Paradox which focuses overmuch on quantification or which overemphasizes certain special cases of the paradox — e.g., the transitive submodel case. As we have seen, Skolem's Paradox comes in many forms, and, even at the technical level, these different forms require different kinds of solutions. This point carries over to the philosophical level as well. It's clear, for instance, that Skolem's Paradox may lead us to ask difficult questions about things like the indefinite extensibility of the concept of set or the coherence of absolute notions of quantification, but I doubt very much that answers to these questions will enable us to provide a (complete) solution to the paradox itself. At best, they will help us to solve those instances of the paradox which most clearly turn on the interpretation of quantifiers. For philosophy, then, as much as for mathematics, a full solution to Skolem's Paradox will have to focus on the fine-grained analysis of *many different* models of set theory (where these different models give rise to different philosophical questions).

These, then, are what I take to be the main points of the preceding discussion. I want to close by considering a worry which this entire discussion may have provoked. So far, I have treated Skolem's Paradox as though it were an essentially *technical* matter. I started by taking both Cantor's theorem and the Löwenheim-Skolem theorem at face value — by, that is, taking a naively realistic attitude towards the mathematics lying behind these two theorems — and I then tried to explain why, understood in this manner, the theorems don't conflict with each other. In doing so, I felt perfectly free to make use of expressions like "the ordinary English understanding of membership," "the real members of $\hat{m}$," "quantifiers which range over the *whole* set-theoretic universe," etc.

The worry here is that this analysis might be a bit *too* naive. At the most basic level, this worry flows from simple incredulity at the idea that anything as strong as full classical set theory can simply be *presupposed* when solving puzzles like Skolem's Paradox.[54] Moving deeper, our incredulity can be reinforced by recalling

---

the context of specific models.

[54] In the literature, incredulity about appeals to our "ordinary English" understanding of set theory is often expressed in terms of opposition to "Platonism." See [Fine, 1968], [Klenk, 1976], [Thomas, 1968], and [Thomas, 1971] for some examples of this way of putting things.

other philosophical puzzles about the interpretation of mathematical language — i.e., puzzles which call into question the determinacy of naive talk about things like "membership" or "the whole universe of sets."[55] Finally, the entire development of twentieth-century set theory may seem to tell against my approach to Skolem's Paradox. After all, the standard response to the classical paradoxes has been to move *away* from naive approaches to set theory and *toward* formal axiom systems (and especially first-order axiom systems).

Clearly, addressing these kinds of worries in any detail would lead us rapidly into deep waters. I don't plan to do that here. Indeed, I won't even try to develop these worries more completely or to track them through the relevant literature. Instead, I'll just make a few short remarks in defense of the kind of technical solution given in sections 2–5. For obvious reasons, I don't regard these remarks as a complete response to the above worries; at best, they constitute a gesture in the direction of such a response.

Let's begin by recalling the *point* of Skolem's Paradox. In theory, the paradox highlights a certain incoherence — or perhaps even an inconsistency — in our ordinary ways of thinking about set theory. More specifically, it purports to show that there is a conflict between the naive acceptance of Cantor's theorem and certain instances of the Löwenheim-Skolem theorem. Since the Löwenheim-Skolem theorem is, presumably, unassailable, this leads to the conclusion that Cantor's theorem should not be taken at face value — i.e., that we should view naive talk about "absolutely uncountable sets" as problematic and to be avoided.

Notice the order of argument here. We start with a naive acceptance of Cantor's theorem. (At the very least, we start with an open mind towards this theorem and towards the naive set theory which lies behind it.) We then formulate Skolem's Paradox. The paradox shows that there is a problem with our initial naiveté, and it forces us to abandon our original acceptance of "ordinary-English" formulations of set theory. In short: Skolem's Paradox does the philosophical work here, and the problematization of ordinary-English set theory is (part of) the philosophical payoff.

My concern, then, is that the worries we're now discussing effectively reverse this order of argument. They *start* with a rejection of ordinary-English set theory — start, that is, with the very thing that Skolem's Paradox is supposed to help us establish — and they then *use* this rejection as a means of defending Skolem's Paradox against certain technical challenges (e.g., those in sections 2–5). On this way of proceeding, however, it's hard to see how Skolem's Paradox still does any real philosophical work. On the surface, it's our initial worries — and whatever arguments may lie behind them — that do the real philosophical work; Skolem's Paradox just tags along for the ride. Let me put this point another way. Anyone who comes to set theory with serious worries about the determinacy (or even the coherence) of ordinary talk about sets and membership will, *of course,* have corresponding worries about the solution to Skolem's Paradox which I developed in sections 2–5. But, they will also have *independent* worries about the notions of countability and uncountability (since these notions are, after all, defined in terms of the problematic notions of membership and quantification over the set-theoretic universe). As a result, there's no need for them to bring Skolem's Paradox into the discussion. Given their initial worries, they have direct arguments against naive talk about "absolutely uncountable sets," and Skolem's Paradox becomes completely superfluous. In short: taking these kinds of worries seriously doesn't help to make Skolem's paradox more significant. On the contrary, it threatens to reduce the paradox to a mere technical side show.

---

[55] See [Klenk, 1976] for an attempt to parley one of Benacerraf's classical puzzles — that presented in [Benacerraf, 1965] — into this kind of challenge to naively technical solutions to Skolem's Paradox. See [Wright, 1985] for a similar argument based on Wittgensteinian considerations concerning the relationship between meaning and use.

Here's one more (and final) way of thinking about all this. To solve Skolem's Paradox, we need to show that there's no conflict between Cantor's theorem and the Löwenheim-Skolem theorem. But that's *all* we need to do. We don't *also* have to solve every other problem in the philosophy of set-theory — i.e., we don't have to solve them *before* we can use words like "set" and "membership" to provide a solution to Skolem's Paradox. In saying this, I'm not trying to dismiss these other problems; I'm just emphasizing that they are, in fact, *other problems.*[56] When we focus resolutely on Skolem's Paradox itself — on the purported conflict between Cantor and Löwenheim — then we find that the technical analysis of sections 2–5 is exactly what we need to solve *this particular puzzle.* In the present context, that's all we need to do.

This, then, explains why I'm at least inclined towards a wholesale dismissal of the kinds of worries now under discussion. It's not that I think that these worries are trivial or misguided (some of them clearly aren't); it's just that I don't think that attending to these worries helps us to understand Skolem's Paradox itself (indeed, I think the worries tend to trivialize the paradox). Of course, I'm aware that I'm evading all of the argumentative details here: to really make these thoughts stick, I'd have to develop the above worries in far more depth and to explore their interaction with Skolem's Paradox in far more detail. But that is a project for another time. For now, I'll simply end with a final summary of this paper: there is no conflict between Cantor's theorem and the Löwenheim-Skolem theorem.

---

[56] An analogy from ethics may be helpful here. Presumably, it's legitimate to discuss topics like abortion or just war theory without first solving the Liar Paradox and/or the problem of induction. This is true *even though* we may want to use the word "truth" in our ethical discussions and/or make some empirical generalizations. To think otherwise is to think that we need to solve *all* philosophical problems in order fruitfully discuss *any* of them. *That's* neither a reasonable nor a profitable standard for good philosophical discourse.

BIBLIOGRAPHY

[Bays, 2000] T. Bays. *Reflections on Skolem's Paradox.* PhD thesis, UCLA, 2000.

[Benacerraf, 1965] P. Benacerraf. What the numbers could not be. In *Philosophy of Mathematics*, pages 272–294. Cambridge University Press, Cambridge, 1983.

[Benacerraf, 1985] P. Benacerraf. Skolem and the skeptic. *Proceedings of the Aristotelian Society*, 59:85–115, 1985.

[Ebbinhaus *et al.*, 1994] H. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic.* Springer, Amsterdam, 1994.

[Fine, 1968] A. Fine. Quantification over the real numbers. *Philosophical Studies*, 19:27–31, 1968.

[George, 1985] A. George. Skolem and the Löwenheim-Skolem Theorems. *History and Philosophy of Logic*, 6:75–89, 1985.

[Hart, 1970] W. Hart. Skolem's promises and paradoxes. *The Journal of Philosophy*, 67:98–109, 1970.

[Jech, 1978] T. Jech. *Set Theory.* Academic Press, San Diego, 1978.

[Keisler and Morley, 1968] J. Keisler and M. Morley. Elementary extensions of models of set theory. *Israel Journal of Mathematics*, 6:49–65, 1968.

[Klenk, 1976] V. Klenk. Intended models and the Löwenheim-Skolem Theorem. *The Journal of Philosophical Logic*, 5:475–489, 1976.

[Kunnen, 1980] K. Kunnen. *Set Theory.* North-Holland, Amsterdam, 1980.

[McIntosh, 1979] C. McIntosh. Skolem's criticisms of set theory. *Nous*, 13:313–334, 1979.

[Myhill, 1967] J. Myhill. On the ontological significance of the Löwenheim-Skolem Theorem. In I. Copi and J. Gould, editors, *Contemporary Readings in Logical Theory*, pages 40–51. Macmillan, New York, 1967.

[Putnum, 1980] H. Putnam. Models and reality. In *Realism and Reason*, pages 1–25. Cambridge UP, Cambridge, 1983.

[Resnik, 1966] M. Resnik. On Skolem's Paradox. *The Journal of Philosophy*, 63:425–438, 1966.

[Shoenfield, 1967] J. Shoenfield. *Mathematical Logic.* Association for Symbolic Logic, Natick, MA., 67.

[Skolem, 1967] T. Skolem. Some remarks on axiomitized set theory. In Jean van Heijenoort, editor, *From Frege to Gödel*, pages 290–301. Harvard, Cambridge, Mass., 1967.

[Thomas, 1968] W. Thomas. Platonism and the Skolem Paradox. *Analysis*, 28:193–196, 1968.

[Thomas, 1971] W. Thomas. On behalf of the Skolemite. *Analysis*, 31:177–186, 1971.

[van Dalen, 1997] D. van Dalen. *Logic and Structure.* Springer, Amsterdam, 1997.

[Wright, 1985] C. Wright. Skolem and the skeptic. *Proceedings of the Aristotelian Society*, 59:116–137, 1985.

OBJECTUAL AND SUBSTITUTIONAL INTERPRETATIONS OF THE QUANTIFIERS

Michael Hand

INTRODUCTION

The ordinary language quantification "There's a cat in my kitchen" does not require for its truth that any of the cats in my kitchen have names. Likewise, "Every cat is in my kitchen" is not made true simply by the presence in my kitchen of all the cats with names. Ordinarily $\forall$ and $\exists$ are understood in the same way. For φ with only x free, we can say roughly,

$(\forall x)\varphi$ is true iff every individual satisfies φ;

$(\exists x)\varphi$ is true iff some individual satisfies φ.

This is the *objectual* interpretation of (first-order) universal and existential quantification. The respective truth-theoretic inductive clauses have the following forms (or something very akin to them).

$\mathbf{s}$ satisfies $(\forall x)\varphi$ iff for every $\alpha \in$U, $\mathbf{s}^{(x/\alpha)}$ satisfies φ,

$\boldsymbol{s}$ satisfies $(\exists x)\varphi$ iff for some $\alpha \in$U, $\mathbf{s}^{(x/\alpha)}$ satisfies φ,

where U is the universe of quantification, $\boldsymbol{s}$ an assignment of members of U to variables, and $\boldsymbol{s}^{(x/\alpha)}$ an assignment differing from $\boldsymbol{s}$ at most by assigning α to x.

Note that a truth theory for a language with objectual quantifiers detours through an inductive definition of satisfaction, as seen here. The substitutional interpretation of $\forall$ and $\exists$ does not require this.

$(\forall x)\varphi$ is true iff every instance of φ is true.

$(\exists x)\varphi$ is true iff some instance of φ is true.

where an *instance* of a (closed) quantification $(\forall x)\varphi$ or $(\exists x)\varphi$ is the result of putting an individual constant for the free variable in φ.

It will help henceforth if we reserve the symbols $\forall$ and $\exists$ for objectual quantification and use Π and Σ for their substitutional counterparts. The question whether the quantifiers of a given language are to be construed objectually or substitutionally becomes the question whether that language's universal and existential quantifiers are to be represented as *our* $\forall$ and $\exists$ or rather as Π and Σ. We can use this convention to give the truth-conditions for substitutional quantifications as

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

$$(\Pi x)\varphi \text{ is true iff } (\underline{\forall} t)(\varphi(x/t) \text{ is true}),$$
$$(\Sigma x)\varphi \text{ is true iff } (\underline{\exists} t)(\varphi(x/t) \text{ is true}).$$

where t ranges over individual constants and $\varphi(x/t)$ is the result of putting the name t for all free occurrences of x in φ. (For metalinguistic quantifiers we simply underline. In general, we remain as casual as possible on the use-mention distinction, as the reader may have already noticed.) The right-hand sides are clearly seen to involve objectual metalinguistic quantification. Philosophical interest in the distinction between objectual and substitutional quantification falls for the most part into two sorts, ontological and meaning-theoretical, though these are not unrelated. The bulk of what follows is devoted to such philosophical issues, but it will be useful first to make some purely logical observations on the objectual/substitutional distinction.

In general, the truth-condition of $(\Pi x)\varphi$ is just that of a possibly infinitary conjunction of all φ's instances (depending on whether the language has a finite or infinite number of individual constants; if the former, substitutional quantification is redundant anyway). Let $< t_0, t_1, \ldots >$ enumerate the individual constants. Then $(\Pi x)\varphi$ amounts in effect to

$$\varphi(\mathrm{x}/\mathrm{t}_0) \& \varphi(x/t_1) \& \ldots$$

and the truth-condition for $(\Sigma x)\varphi$ is similarly the possibly infinitary disjunction. This observation is merely heuristic, since we assume an ordinary language with only finitely long wffs. (Typically we take only individual constants as substuends; this restriction can be loosened to allow other terms, with certain restrictions imposed.)

If for every member of the domain U of quantification there is a name in the language, then objectual quantifications and their substitutional counterparts coincide in truth-value. If not, however, then $(\forall x)\varphi$ may fail while $(\Pi x)\varphi$ holds, since all the *named* individuals may satisfy φ and thus all instances of φ are true, yet some unnamed individuals may nonetheless fail to satisfy φ. When only *unnamed* individuals satisfy φ, $(\exists x)\varphi$ holds though $(\Sigma x)\varphi$ fails. This is the *problem of too few names*: in general $\forall$ is stronger than Π, and $\exists$ weaker than Σ; only on the assumption that all individuals have names are objectual and substitutional quantifications equivalent. (A wff φ is stronger than a wff ψ iff for any wff χ, if $\psi \vDash \chi$ then $\varphi \vDash \chi$.)

It was mentioned above that although a truth theory for a language with objectual quantification requires an inductive detour through the notion of satisfaction, with truth defined in terms of satisfaction of *closed* wffs, this detour is unnecessary for substitutional quantification. The reason the detour is needed in customary semantic treatments of languages with objectual quantifiers is that the semantic apparatus determines the semantic value of a nonatomic wff on the basis of the semantic values of its *immediate* syntactic constituents (the wffs mentioned on the right-hand sides of the inductive clauses). In the case of quantifiers, this means that the specification of a semantic value of a closed wff requires semantic values to be assigned to *open* wffs. (We reserve "sentence" for *closed* wffs.) The truth-values of *sentences* $(\forall x)\varphi$ and $(\exists x)\varphi$ are determined by the value of φ, which has at most x free. (Some languages allow superfluous quantification, where φ is already closed, but this is beside the present point.) Intuitively, a wff φ that has only x free is not (typically) true or false, but rather true or false *of* something. This is codified in the notion of a wff's satisfaction by an assignment of individuals to variables. Truth of a sentence is then defined as satisfaction by all assignments. In contrast, substitutional quantifiers require only that *sentences* be accorded semantic values, and (closed) quantified sentences are treated not by reference to the semantic values of their immediate syntactic constituents, but instead by means of the semantic values of their *instances.* This feature of substitutional quantification may thus open ontologically interesting semantic possibilities, depending on how the semantic values of quantifier-free sentences, ultimately atomic sentences, are determined.

A quick inspection of the quantifier rules belonging to a perspicuous deductive system (natural deduction trees) shows how $\forall$ is stronger than Π and $\exists$ weaker than Σ in the case where not all members of U have names in the language. In particular, the rule of $\forall$-introduction is too demanding to capture *merely* the meaning of Π, and the $\exists$-*elimination* rule too weak to capture merely that of Σ.

> $\forall-$*introduction*: Conclude $(\forall x)\varphi$ from $\varphi(x/t)$, when $\varphi(x/t)$ depends on no undischarged assumptions containing t.

The effect of the condition for application of this rule is to ensure that the subproof above $\varphi(x/t)$ remains valid on any *reinterpretation* of t. (This rule can be formulated with a parametric free variable instead of a parametric name, and thus serve as a rule of inference for a nameless language.) Thus the condition requires, for the applicability of the rule, not merely that for each name t', every wff $\varphi(x/t')$ be true (on the given assumptions), but rather that $\varphi(x/t)$ be true (on the given assumptions) no matter *which* individual the name t is taken to name, even when the individual is *in fact* nameless in the language. $\exists$-*elimination* is a bit more complicated.

> $\exists-$*elimination*: From the major premise $(\exists x)\varphi$ and the minor premise ψ, conclude ψ and discharge from the subproof above the minor premise the assumption $\varphi(x/t)$, when (i) neither $(\exists x)\varphi$ nor ψ contains t, and (ii) no remaining undischarged assumptions of that subproof contain t.

The effect of this condition is again to ensure that the subproof in question remains valid on any reinterpretation of the parametric name. This, assuming the truth of the existential major premise on *its* given undischarged assumptions and of the remaining undischarged assumptions of the subproof in question, guarantees the drawn conclusion.

Note, however, that these rules remain valid for the substitutional quantifiers. In both cases, the "conditions" on the applicability of $\forall$-*intro* and $\exists$-*elim* are sufficient to validate the substitutional counterparts of the rules. If the proof or subproof in question remains valid on *any* reinterpretation of the involved name, then it remains valid on any reinterpretation of the name over just the *named* individuals.

This is related to the fact that the semantic consequence relation (entailment) is compact for first-order objectual quantification but not for substitutional. For a set of sentences Γ and a sentence φ, $\Gamma \vDash \varphi$ when any interpretation of the language whereon all members of Γ are true is such that φ is also true on that interpretation. (We assume the language has an infinite set of individual constants. An *interpretation* of a first-order language consists in a nonempty universe U — a set of individuals — with denotations from U for all individual constants, n-ary relations on U for all the n-adic predicates, and n-ary functions on U for all the n-adic function symbols of the language.) Compactness of $\vDash$ is the following property.

> If $\Gamma \vDash \varphi$ then there is a finite $\Gamma' \subseteq \Gamma$ such that $\Gamma' \vDash \varphi$.

Now observe that

$$\{\varphi(x/t_0), \varphi(x/t_1), ...\} \vDash (\Pi x)\varphi$$

although no finite subset of this infinite one entails $(\Pi x)\varphi$. Because some universes have infinite U but not every member of U has a name in the language, even the infinite set does not entail $(\forall x)\varphi$. (Thus ω-inconsistency is intrinsically an objectual notion.)

Up to this point we have been talking specifically about substitutional quantifiers that involve substitution of individual constants for variables, but substitutional quantification need not be so limited. In general, the semantics of substitutional quantification requires only a specified substitution-class of expressions available to serve as substituends into the sites occupied by a wff's variables, yielding the sentences whose truth or falsity determine the truth-value of the substitutionally quantified sentence. This substitution-class might consist in, for instance, predicate constants of the right adicity, so that "higher-order" substitutional quantification is possible, or it might consist in sentences themselves, or even punctuation marks of the language, e.g., parentheses.

QUINE AND ONTOLOGICAL COMMITMENT

W. V. Quine is famous for his "criterion of ontological commitment". Let a theory T in a given language be closed under entailment. We can formulate Quine's idea as follows.

> When a theory T includes an existential objectual quantification $(\exists x)\varphi$, it is ontologically committed to (i.e., committed to the existence of) an individual satisfying φ.

This criterion can be formulated with only the existential quantifier because $(\forall x)\varphi$ entails $(\exists x)\varphi$. (It is important that Quine also inveighs against higher-order quantification, for independent reasons, and assumes that T does not contain higher-order quantifiers.) Alternatively put, T is committed to the existence of all the members of the universe U of quantification, the values assigned to variables. Thus the well-known Quinean dictum, "to be is to be the value of a variable".

Substitutional quantification lacks such ontological potency. Its semantics involves *metalinguistic objectual* quantification over *substituends*, expressions of the object language, not over members of a universe of object-language reference. Substitutional quantification is therefore inadequate for such ontological work as Quine has in mind with his criterion. As he puts it [1969, 107],

> Substitutional quantification in the substitution class of singular terms, or names, is the sort that comes closest to objectual quantification. But it is clearly not equivalent to it — not unless each of our objects is specifiable by some singular term or other in our language, and no term of that substitution class fails to specify an object. For this reason substitutional quantification gives no acceptable version of existence properly so-called, not if objectual quantification does. Moreover, substitutional quantification makes good sense, explicable in terms of truth and substitution, no matter what substitution class we take — even that whose sole member is the left-hand parenthesis. To conclude that entities are being assumed that trivially, and that far out, is simply to drop ontological questions.

Quine also rejects the thought that *names* are vehicles of ontological commitment. He advocates elimination of names in favor of definite descriptions, and elimination of these in favor of objectual quantification, to produce a properly regimented language within which a theory wears its ontological commitment on its sleeve, in the form of its existential objectual quantifications.

Names are eliminable as follows. For each name t, introduce a monadic predicate N_t true uniquely of t's referent, and convert φ into $\varphi(t/\iota x N_t x)$. The latter can be treated in the familiar Russellian way (the theory of descriptions). The result is a name-free, description-free T. Besides, as Quine mentions, we sometimes use names in ways that do not match up with any (objectual) existential claims of the regimented theory: so-called "empty names," as we shall see. Only names that do match up carry ontological weight, and this is seen in the existential products of their elimination in favor of existential quantifications. It is *these* that indicate ontological commitment. Names cannot be relied upon as indicators of such commitment.

Some have urged that the semantical treatment of substitutional quantification shows only that such quantification merely *defers* the question of ontological commitment, placing the metaphysical weight upon the quantifier-free wffs of the language, especially the atoms. Assuming some prior treatment of empty names distinguishing them from successfully referential ones, and bearing in mind that (first-order) substitutional quantification can be present only for a language from which names have not been regimented away, we might hold *contra* Quine that if the closed atoms are treated by means of a semantics of satisfaction, then an ontological commitment to the members of a domain of values of variables is already present. Thus Dummett [1973, 528]:

> Quine supposes that only quantification over a range of entities requires us to ascribe to the corresponding constant expressions reference to those entities, or commits us to including those entities in our ontology: substitutional quantification thus appears to him as a threat, a device for employing quantification without ontological commitment. The advocates of substitutional quantification share with Quine the assumption that ontological commitment is required only by [objectual] quantification, and therefore regard substitutional quantification as a means of achieving liberation from ontological commitment. Both are wrong, because it is not quantification which in the first place requires the ascription of reference: reference must be ascribed to any expressions which function as significant units of sentences of a language, if we are to be able to frame a semantic account of that language.

REFERENTIAL AND NONREFERENTIAL BASE SEMANTICS

We distinguish in the semantic apparatus two parts: the *base* semantics, which treats atoms, and the *structural* semantics, which treats nonatoms. Structural semantics for a first-order language with objectual quantification and the familiar connectives includes such inductive clauses for satisfaction as those given earlier for $\forall$ and $\exists$, as well as clauses for connectives, e.g.,

$$\boldsymbol{s} \text{ satisfies } (\varphi \& \psi) \text{ iff } (\boldsymbol{s} \text{ satisfies } \varphi \text{ and } \boldsymbol{s} \text{ satisfies } \psi),$$

$$\boldsymbol{s} \text{ satisfies } \sim \varphi \text{ iff not } (\boldsymbol{s} \text{ satisfies } \varphi).$$

The *base* semantics for a language containing objectual quantification contains such clauses as

$$\boldsymbol{s} \text{ satisfies } Px \text{ iff } \underline{P}\boldsymbol{s}(x),$$

where Px is atomic, $\underline{P}$ is the metalinguistic predicate interpreting the object-language predicate P, and $\boldsymbol{s}(x)$ the individual that $\boldsymbol{s}$ assigns to x. Similarly, when t is an individual constant we have

$$\boldsymbol{s} \text{ satisfies } Pt \text{ iff } \underline{P}\boldsymbol{s}(t),$$

with the proviso that assignments do not vary with respect to individual constants.

(It is clear at this point that our semantics cannot serve the purposes of free logic, wherein nondenoting names are permitted.) Assignments assign to both names and variables members of the universe, and thus we call such a base semantics "referential".

A base semantics for a language with substitutional quantification (but lacking objectual quantification) has no more need for the notion of satisfaction than does the structural semantics. It need only treat *closed* atoms, and may avoid the whole semantical machinery of assignments. It becomes possible to invoke some nonreferential notion of "truth" for closed atoms. In fact, it is inappropriate to describe substitutional quantifiers as "ranging" over "values" for their variables at all. If, however, the base semantics does involve satisfaction, then the language already involves ontological commitment by providing a universe whose members are the values of assignments. Indeed, even if objectual quantifiers are absent from the language, such a referential semantics is already ontologically committed. Some have argued that, contrary to the idea underlying Quine's idealization of a completely regimented, quantifier-free name-free language of a theory, it is names in particular that are basic to objectual reference. As Neil Tennant [1979, 40] puts the case in a nutshell,

> What makes [names] basic to the semantic enterprise is this: no language could be taken as speaking of objects without our grasping what it is (or would be) to refer to an object by name. Even a language "with [objectual] quantifiers but without names" can have its semantic workings made clear only by drawing on an implicit understanding of the name-bearer relation. The semantics of [objectual] quantification requires at the very least a notion of assignment of individuals to variables. This is the very same notion as that of temporarily treating variables as names.

If such comments are correct, we must qualify Quine's criterion of ontological commitment. Of a language interpreted by means of referential base semantics we might say that it "has room" for names and for objectual quantification even if in fact it has neither of these in its vocabulary: the base semantics relies upon the idea of assigning objects to variables, and the language's namelessness indicates only that there are no expressions concerning which it is required that assignments do not vary with respect to them. When objectual quantification is present, the only *overt* object-language indication of ontological commitment is made by them. Still, our ontological commitment stems from the referential base semantics whereby members of the quantificational universe are assigned to variables. Objectual quantification is important because it *requires* such treatment. If the language is quantifier-free, nothing in it *necessitates* that a theory be ontologically committed, but we must see it as carrying such commitment if we take its atoms to be referential. Interpretation of T's language by means of a referential base semantics — even in the absence of objectual quantification — suffices on its own to carry ontological commitment. Quine's point may thus be taken as the claim that unless objectual quantification is present, T is subject to a nonreferential semantical treatment. Perhaps Quine would prefer this point to be formulated as follows: absent objectual quantification, T itself is ontologically noncommittal in that proponents of T do not *ipso facto* commit themselves to entities that T

appears (intuitively, given our referential habits of interpretation) to be about; to secure their freedom, they must hold that this is *mere* appearance.

THE "SUBSTITUTIONAL STRATEGY"

The suggestion that substitutional quantification may open new ontological possibilities and thus alleviate some metaphysical worries has impressed some theorists. We shall examine a few uses of the "substitutional strategy" for avoiding these worries. To address the first example, we need to remind ourselves of Quine's "three grades of modal involvement" and his special antagonism toward the third grade. Quine is no fan of the analytic/synthetic distinction, nor of modal logic. In particular, however, he objects to the third grade, "quantifying in," on distinctive grounds.

The first grade of modal involvement consists in application of a metalinguistic modal predicate to metalinguistic names of sentences, as in

$$Nec(\text{“}\varphi\text{”}),$$

where the metalinguistic predicate expresses some necessity-like property of sentences, e.g. logical validity. Quine of course has no objections to reading the *Nec* predicate as "is logically valid". To this extent, our involvement with this "modal" predicate and its dual, *Poss*, is no more problematic than our commitment to the notion of logical validity, though a risk is involved if we try to widen this notion of necessity to include such "necessary" sentences as "All bachelors are unmarried," i.e., to apply it to analytic claims that are not logically valid. Still, use of the restricted necessity predicate is intrinsically unobjectionable, and no more objectionable that this is the move to the second grade, where we introduce into the object language a "statement operator" $\Box$ applicable only to sentences (closed wffs). Stipulating for a closed φ that $\Box\varphi$ is true iff φ is logically valid is again risky, but not itself objectionable on semantic grounds. The danger of this second grade lies in the fact that attempts to encode in the object language a device for expressing a feature of φ whose attribution should properly occur in the metalanguage invites troublesome (but avoidable) further moves. Historically, this second grade of modal involvement is, in virtue of the attempt to capture in the object language a properly metalinguistic predication is "conceived in sin": failure to observe the distinction between use and mention. It invites pointless research into iterated modalities, losing sight of the fact that $\Box\Box\varphi$ should, on the metalinguistic interpretation of $\Box$ as *Nec*, involve us in yet a higher metalanguage: $Nec(\text{“}Nec(\text{“}\varphi\text{”})\text{”})$. But we can take care, and so our involvement in the second grade is not *ipso facto* problematic.

An important further danger of the second grade of involvement is that it invites an extended use of $\Box$ not merely as a statement operator but as a "sentence operator" (what we would call a *wff* operator), attachable to an *open* wff, the result of which can then be closed by a quantifier, for instance $(\exists x)\Box Fx$. This is "quantifying in". $\Box$ is no longer understandable in terms of a metalinguistic predication, but requires for its interpretation something radically different. Logical validity is at least a property of *sentences*, but quantifying in cannot be understood in terms of a metalinguistic attribution of any property of sentences. Rather, the third grade requires that sense be made of the *open* wff $\Box Fx$, and this seems to require attribution of a modal property to an individual independently of any particular means of referring to it. For example, we might (contentiously) hold

(1) Nine is odd

to be logically true, thus necessary, and so endorse

(1′) $\Box odd(9)$.

(For the moment we abandon our restriction to languages whose only terms are names.) We might (uncontentiously) also take

(2) The number of planets is odd

to be *not* logically true, and thus endorse (with P for "numbers the planets")

(2′) $\sim\Box odd(\iota x Px)$

But nine *is* the number of planets, so there is already a new peculiarity present here at the second grade. This unfortunate state of affairs is explainable on the grounds that replacement in a logically valid sentence of one denoting phrase for a coreferential different one cannot be expected to preserve validity. Yet we still cannot make sense of (3). (*Whatever* we take *Nec* to mean, no matter how ill-conceived, e.g., analyticity, we create new problems when we move to the third grade.)

(3) $(\exists x)\Box odd(x)$

This existential generalization from (1′) forces us to the view that there is some number (nine? the number of planets?) satisfying $\Box odd(x)$. That *it* satisfies $\Box odd(x)$ cannot be explained in terms of metalinguistic ascriptions of properties to any *sentence*, and so it appears that we must invoke the idea of a number's being necessarily odd *independently* of how it is referred to. Quine concludes that the third grade of modal involvement forces "Aristotelian essentialism" upon us. We may distinguish the third grade by attributing to the second grade of modal involvement the view that necessity is *relational*: an object may possess a property necessarily, relative to one way of referring to it, but not necessarily, relative to another way. The number nine is necessarily odd relative to the expression "9", but at the same time it is not necessarily odd relative to the coreferential expression "the number of planets". The third grade removes this relativity, and with it any hope of treating necessity as a linguistic phenomenon at all.

Reinterpretation of quantifiers may seem to remove this worry. Instead of treating $(\exists x)\Box Fx$ as an objectual quantification, we may treat it *substitutionally*, i.e., we read it as $(\Sigma x)\Box Fx$. Its truth does not require us to assert that some *individual* is necessarily F, but only that for some term t of the language, $\Box Ft$ is true. This suggestion has immediate implications for Quine's worry over the third grade of modal involvement because it preserves the linguistic nature of necessity in the face of quantifying in.

Identity in a language with opacity, such as the one at hand here, is a very delicate business. The principle of substitutivity of identity is an objectual principle,

$$(\forall x)(\forall y)(x = y \rightarrow (\varphi \rightarrow \varphi')),$$

where φ' results from φ by replacing any number of free occurrences of x therein by y. The corresponding substitutional principle $(\Pi x)(\Pi y)(x = y \quad \rightarrow (\varphi \rightarrow \varphi'))$ may fail, as above. (For this reason we subsequently focus mostly on substitutional languages lacking identity.) Thus substitutional reconstrual may motivate the imposition of restrictions on the substitution-class of terms associated with the quantifiers (cf. Marcus).

Another potentially useful employment of the substitutional strategy is its application to such examples as (5), which we might wish to endorse in view of our willingness to hold (4) true.

(4) Pegasus is a flying horse.

(5) There is a flying horse.

The Russell/Quine elimination of names results in the verdict that (4) is false, for it takes (4) already to have the logical form of an objectual quantification. If the covert existential quantifier of (4) and the overt one of (5) are construed as objectual, then both (4) and (5) must be denied. Even if the Russell/Quine way of regimenting (4) is rejected and (4) is allowed to stand at face value, it entails (5) and so must be counted as false anyway. Of course, a desire to regard (4) as being true cannot be accommodated by means of a straightforward referential base semantics, since the atoms "Pegasus is a horse" and "Pegasus flies", of which (4) can be treated as a conjunction, must thereby be interpreted by means of assignments that (uniformly) assign a member of the universe to "Pegasus" and presumably have disjoint extensions of the predicates "is a horse" and "flies" (lest we be ontologically committed to at least one flying horse, Pegasus). But if some other semantical story is told about (4), so that (4) turns out true, a substitutional construal of quantifiers may seem happy: (5), so treated, is true in virtue of (4)'s truth: (Σx) (x is a horse & x flies).

In a related way, substitutional quantification might be thought to avoid certain problems in the semantics of propositional attitude constructions. While we may assume that

John believes that Hesperus is a planet

is true, and given the truth of

Hesperus = Phosphorus,

it nonetheless seems that

> John does not believe that Phosphorus is a planet

may also be true. This problem especially afflicts the view that proper names are directly referential, having no descriptive content and therefore no difference to be appealed to in accounting for the failure of substitutivity here. We appear to be committed to

> $(\exists x)(\exists y)(x = y$ and John believes that x is a planet and John does not believe that y is a planet),

which yields a contradiction,

> $(\exists x)$(John believes that x is a planet and John does not believe that x is a planet).

This distasteful conclusion is avoided when the quantifiers are construed substitutionally. The double quantification is true, but there is no entailment to the single quantification.

This strategy of substitutional reconstrual seems to have potential for avoiding other metaphysical problems. In particular, it may seem attractive to *nominalists.* Nominalism is traditionally understood as an ontological view that denies existence to anything other than individuals. Thus a traditional nominalist will object to the higher-order quantification of (6) when regimented as (6′).

(6) There is something that Alicia and Beth both are.

(6′) $(\exists X)(Xa \& Xb)$

A second-order objectual quantifier as in (6′) is treated as follows. For monadic predicate variables, the universe of quantification U_2 is the power set (or nonstandardly, a proper subset of the power set satisfying certain conditions) of the first-order universe U_1 (the set of individuals); for n-adic predicate variables, the universe is the n-th Cartesian product U_2^n, the set of n-tuples of members of U_1 (or nonstandardly, a proper subset of it satisfying certain conditions). An assignment now treats second-order variables and constants as well, assigning them members of the appropriate universe and always taking the same value for second-order constants, which in the monadic case are thus treated as "names" of sets. This ontological commitment to sets (or, on a more fine-grained treatment, to properties), as Quine would have it, is specifically due to the presence of the *objectual* higher-order quantifier in (6′).

Let Alicia and Beth be police officers. Unless we are prepared ontologically to admit the *set* of police officers (not to mention the *property* of being a police officer), (6′) should trouble us. As nominalists we may respond by rejecting higher-order objectual quantification. What, then, to do with (6)?

On the other hand, (6″) would seem to avoid this metaphysical trouble.

(6″) $(\Sigma X)(Xa \& Xb)$

(6″) is true iff for some substituend of the appropriate syntactic category (monadic second-order constants), the corresponding substitution-instance of $Xa \& Xb$ holds. Indeed, (6″) is true because (7) is true.

(7) $\mathit{police\text{-}officer}(a) \& \mathit{police\text{-}officer}(b)$

Even if, in our base semantics, we interpret the predicate by means of a set, (7) is first-order and does not require sets as *values of variables.* The unreconstructed Quine, and with him our nominalists, will breathe a sigh of relief.

Propositions are simply zero-place properties, and so an ontological distaste for properties should apply to propositions. Yet from our earlier

John believes that Hesperus is a planet

we may conclude

$(\exists p)$(John believes p).

Here the quantification appears to range over propositions. Substitutional quantification yields the supposedly unobjectionable

(Σp)(John believes p).

Another allegedly ontologically potent use of substitutional quantification concerns arithmetic (cf. Gottlieb 1980). A philosopher who wishes to avoid ontological commitment to the natural numbers may choose to construe the quantifiers of first-order arithmetic as substitutional.

There is a prime between 7 and 12

may be read as

$(\Sigma x)(\mathit{prime}(x) \ \& \ (7 < x \ \& \ x < 12)$

and thus is true in virtue of the truth of

$\mathit{prime}(11) \ \& \ (7 < 11 \ \& \ 11 < 12)$.

Of course the defined predicate *prime* involves quantification as well, which will again be construed substitutionally. (Fortunately for the substitutional strategist, all the natural numbers have names.) The general idea of this approach to avoiding ontological commitment is obvious. Given a discourse with respect to which we wish to avoid the ontological commitments of objectual quantification, invoke substitutional quantification instead (but a nonreferential base semantics must also be provided, if the earlier remarks on Quine's criterion of ontological commitment are correct).

Set theory provides another vivid, and ultimately perhaps less objectionable, example of this ontological strategy. It is well known that there is an intimate connection between Zermelo-Fraenkel set theory with choice (ZFC) and Von Neumann-Bernays-Godel set theory (NBG). Although ZFC is strictly a *set* theory, NBG permits *proper classes*: set-like objects — "collections" — that are not sets. Proper classes cannot be members, so the set-theoretic paradoxes (e.g. Russell's) are avoided. In particular, the proper classes of NBG are "predicative": although NBG has comprehension axioms asserting the existence of classes in general (and distinguishes sets as those classes not equinumerous with the set-theoretic universe, the so-called limitation of size principle), all comprehension axioms are formulated with quantifiers, other than the initial existential, ranging only over sets (and the axioms of ZFC that are not already provable for sets in NBG are restricted to sets as well). For φ with at most x free, and all quantifiers in φ restricted to sets,

$$(\exists Y)(\forall x)(x \in Y \text{ iff } \varphi).$$

(These axioms are typically formulated with lower-case set-variables and upper-case class-variables.) So the only *proper* classes admitted are those answering to these *restricted* comprehension axioms.

The intimate connection mentioned above is this: NBG is conservative over ZFC. Roughly, NBG tells us nothing more about sets and how they relate to other *sets* than ZFC already tells us. The metaphysical question is then whether ZFC has a lesser ontological commitment than NBG. *Prima facie*, and on Quinean grounds, it would seem so, since NBG asserts the existence of things, in fact a strange new *kind* of things, that ZFC does not: proper classes.

If there are, however, conceivable reasons to desire the ontological admission of a mild variety of proper classes, then NBG may seem preferable. Here is a very simple (though fictitious) example. Although the (proper) class of ordinals is not accommodated by the iterative conception of sets embodied by ZFC, we can imagine a "procedure" for collecting them. Each time a new rank of sets is "formed" (by power set or union), add the "new" ordinal rank to our ever-increasing collection of sets' ranks. The peculiarity of this new collection is that it is never completed (though at each stage of the collecting, all it has in it are sets). If we grant that such a "procedure" is sufficient to yield a conception of its totality, we have thus admitted a proper class.

A proponent of a strictly *set* ontology, wishing to avoid this further commitment, has an substitution-strategic option. She may suggest that the initial existential quantifiers of NBG's comprehension principles be construed substitutionally. (Objectual interpretation of set-quantification will re-establish commitment to sets.) In this way, she holds, she avoids ontological commitment to anything other than sets, while permitting a certain amount of ontologically innocuous theorizing seemingly about nonset collections. Indeed, it is not clear that this strategy would have great philosophical advantages over the mere denial that that there are nonset collections, but at least it would not interfere with the advocate's use of ZFC to characterize sets, since NBG is conservative over ZFC, and it accommodates the views of ontologically mild proponents of proper classes, who want only the "predicative" ones of NBG. (Abandoning the predicativity of NBG's comprehension principles results in loss of conservativeness.) But here again, in light of earlier discussion, it is not obvious that this substitutional move has ontological consequences. If the unreconstructed Quine is right, perhaps so. Charles Parsons [1983, 66] writes,

> The advantage of substitutional quantification in [such a] case is that it fits the idea that the classes involved are not "real" independently of the expressions for them...The fact that the substitution interpretation yields truth-conditions for quantified sentences means that everything necessary for speaking of these classes as entities is present, and the request for some more absolute verification of their existence seems senseless.

There remains a further difficulty with the substitutional strategy in general. We've seen a variety of cases where opponents of various ontologies, belonging to various discourses, might invoke substitutional quantification in hopes of avoiding an unpleasant commitment. The worry arises that the substitutional strategy is used *promiscuously*, invoked helter-skelter against whatever ontological threats one wishes to avoid cheaply. Paul Benacerraf [1983] has addressed the general concern of which this — substitutional reconstrual as as a strategy for escape from ontological commitment — is a special instance. He considers (p. 405) specifically the case of arithmetic, and asks whether (7) and (8) should be both construed as having the form (9).

(7) There are at least three large cities older than New York.

(8) There are at least three perfect numbers greater than 17.

(9) There are at least three FGs that bear R to a.

Benacerraf objects to the tendency on the part of some philosophers to posit new construals of arithmetical claims in order to serve some philosophical goal. He proposes a (no doubt defeasible) methodological condition on theories of meaning for different discourses: that their semantics cohere across discourses. Specifically, few would deny that (9) is the logical form of (7), with objectual readings of the quantifiers. Yet some would deny that (9) is the logical form of (8). The primary villains here are philosophers of mathematics who advocate "combinatorial" reinterpretations of arithmetical claims, but his concern applies to proponents of the substitutional strategy as well. He requires a uniform semantical treatment, and writes [p. 408]

> Perhaps the applicability of this requirement to the present case amounts only to a plea that the semantical apparatus of mathematics be seen as part and parcel of that of the natural language in which it is done, and thus that whatever *semantical* account we are inclined to give of

> names, or, more generally, of singular terms, predicates, and quantifiers in the mother tongue include those parts of the mother tongue which we classify as mathematese.

Benacerraf's appeal for semantical uniformity (amounting to a requirement of *referentiality*) can be read in our case as focusing on base semantics. Without a referential base semantics for arithmetical discourse, there is no guarantee that *truth* is what the semantics is assigning to sentences. Objectual quantification guarantees base referentiality, but as we have seen substitutional quantification does not constitute an immediate rejection of it. Substitutional quantification as a means of avoiding ontology is likely to be accompanied by a nonreferential base semantics, for substitutional construal of quantifiers merely pushes the ontological issues downward (or, again as we've seen, so some hold). Substitutionalism is perhaps not itself a particularly egregious violation of Benacerraf's principle, but an advocate who also prefers a nonreferential base semantics must give a lengthy metasemantical account of *truth* to excuse her base nonreferentiality in order to avoid deeply violating the principle.

Another aspect of Benacerraf's case, one that he does not particularly emphasize in the passage above but which deserves attention, is that *discourses interpenetrate.* Consider the substitutional strategy in connection with the third grade of modal involvement. Sentences of the third grade do not belong to a well-insulated discourse. Their quantifiers are, *prima facie*, the quantifiers of much nonmodal language. The consequences of adopting a substitutional construal with respect to them must therefore extend far into the surrounding language. The same is true for the quantifiers of arithmetic. The substitutional strategy applied here can be expected to affect our semantics of a good deal of applied mathematics, and may cause ripples far beyond the limits even of that. On the other hand, the substitutional treatment of the class comprehension axioms of NBG seems considerably less threatening, given an objectual treatment of set quantifiers. No doubt each use of the strategy must be judged on its own semantical and ontological merits — the uniformity requirement is presumably defeasible — but Benacerraf's meaning-theoretical concern is a significant one for substitutionalism in general.

KRIPKE ON SUBSTITUTIONAL QUANTIFICATION

Saul Kripke [1976] is the most important discussion of substitutional quantification since Quine. He answers various meaning-theoretical objections to substitutional quantification, gives a thorough account of the semantics of such quantification, and provides deep insights on the question of ontological commitment, the significance of Tarski's "Convention T", the question of homophonic T-sentences, and other matters. We cannot go into all the contributions of this paper, but will discuss some central ones, beginning with his semantics.

We distinguish notationally between objectual variables, which are bound by objectual quantifiers and whose semantics involves assignments to them of individuals, and substitutional variables, marking the sites of substitution by expressions belonging to a specified substitution-class and bound by substitutional quantifiers. Use italicized "x_1", "x_2", etc. for the former and roman "x_1", "x_2", etc., for the latter. (We continue to write simply "x", "y", "x" and "y" for convenience.)

Given a language L and a semantics for it — call the L-sentences *atoms* regardless of their own internal structure — we extend L to L' by introducing the usual connectives and also $(\Sigma \mathrm{x})$. (Kripke introduces $(\Pi \mathrm{x})$ by definition as the dual of $(\Sigma \mathrm{x})$, just as he has only & and $\sim$ for primitive connectives, but this is inessential.) The logical vocabulary introduced in this extension must be notationally different from the logical vocabulary of L. The L'-connectives are interpreted customarily. Let C (the substitution-class) be a set of L-expressions; call the members of C *terms*. Kripke allows that C need not be a set of individual constants or higher-order constants, but merely a set of *expressions* of L, perhaps including punctuation symbols or even "ungrammatical" L-expressions, e.g., those consisting of L-sentences followed by binary L-connectives. An atomic *preformula* of L' is the result of replacing terms in an L-sentence by substitutional variables. When an atomic preformula is such that replacement of its variables by any terms yields back an L-sentence, it is an atomic *formula* of L$'$. (This ensures that C is not *motley*, as we will use the term later. All members of C are *grammatically available* substituends for any variable bound by the associated quantifier, even though the L-expressions belonging to C need not themselves be grammatical.) In Lesniewski's extreme case where C contains only the left-hand parenthesis,

$$\mathrm{x}\varphi\&\psi)$$

is a formula when φ and ψ are L-sentences and the ampersand is L's conjunction-connective (and the parentheses its punctuation devices), because replacement of x therein by any member of C yields back an L-sentence, $(\varphi\&\psi)$. Note that Kripke's treatment is considerably more general than those we have discussed so far: he calls the L-sentences "atomic", but in fact they may already involve connectives and quantifiers (but not the new ones of the extension). The formulas of L' are defined inductively on the basis of "atomic" L'-formulas in the usual way. Truth (not satisfaction) is inductively defined over *closed* conjunctions, negations, etc., of L'-formulas. Where φ has at most x free, $(\Sigma \mathrm{x})\varphi$ is true iff there is a $t \in$C such that $\varphi(\mathrm{x}/t)$ is true.

However truth is handled in the semantics for L, all that *this* semantics for L'-sentences presupposes is the set $Tr(L)$ of L-truths. The given semantics then suffices to extend $Tr(L)$ uniquely to a set $Tr(L')$ of L'-truths. For instance, when φ is an L'-atomic formula with at most x and y free, $(\Sigma x)(\Sigma y)\varphi$ is true iff there are t, $t' \in C$ such that $\varphi(y/t')(x/t) \in Tr(L)$. This confirms that substitutional quantification is not itself ontologically committal: the commitments of a theory expressed in L' are no more than those of L, *mod* the (objectual) metalanguage's additional ontology of new L'-expressions.

In our earlier terms, the ontological commitments of a substitutional language L' reduce to those of the "base semantics" that determines $Tr(L)$. If the base semantics is nonreferential, and involves only metalinguistic reference to L-expressions, then the additional ontology of L' is null (save for those new L'-expressions). Kripke gives an example of this. Let L contain no logically compound wffs. Let the only L-predicate be the binary predicate M, and the terms of L be finite nonzero strings of the two letters a and b. So all L-sentences have the form $M(t, t')$, where t and t' are L-terms. Let $M(t, t') \in Tr(L)$ iff t' is the mirror-image of t. Then among the L-truths are $M(a, a)$, $M(abb,bba)$, and $M(bbaaba,abaabb)$; L-falsehoods include $M(a,b)$, $M(aba,bb)$, etc. This nonreferential base semantics requires no ontological commitment beyond L-expressions, which is presumably required of any semantics, so we may say loosely that its ontological commitment is null. Extending L to the substitutional L' as above, thus providing the necessary structural semantics accordingly, yields a substitutional semantics that is again null.

Kripke indicates how a language with *both* objectual and substitutional quantification can be handled semantically. Since the former kind of quantification is present, truth in general must be defined in terms of satisfaction, but the recursion clauses for satisfaction will not deal with wffs having free *substitutional* variables. This generalizes the fact that substitutional quantifiers of a language (without objectual ones) allow a direct inductive definition of truth. The set C of available substituends must also be specified. Our base semantics treats "substitutionally closed" atomic formulas. Our inductive clauses are as usual for connectives and $\exists$, but involve satisfaction only of substitutionally closed wffs, so assume that φ has among its *substitutional* variables only x free. The clause for Σ is predictably

$$\boldsymbol{s} \text{ satisfies } (\Sigma \mathrm{x})\varphi \text{ iff } (\underline{\exists} t \in \mathrm{C})(\boldsymbol{s} \text{ satisfies } \varphi(\mathrm{x}/t)).$$

The earlier substitutional extension L' of an "atomic" L (which, recall, need not be a language of mere *atoms* but may even contain objectual quantification and notationally distinct substitutional quantification) does not require clauses like this one. L'-sentences do not have the substitutional quantifiers of L' occurring within the scopes of objectual ones. The present semantics handles such sentences, e.g. $(\forall x)(\Sigma x)\varphi$.

$$\boldsymbol{s} \text{ satisfies } (\forall x)(\Sigma \mathrm{x})\varphi \text{ iff } (\underline{\forall} \alpha \in \mathrm{U})[\boldsymbol{s}^{(x/\alpha)} \text{ satisfies } (\Sigma \mathrm{x})\varphi],$$

so

$$\boldsymbol{s} \text{ satisfies } (\forall x)(\Sigma \mathrm{x})\varphi \text{ iff } (\underline{\forall} \alpha \in \mathrm{U})(\underline{\exists} t \in \mathrm{C})[\boldsymbol{s}^{(x/\alpha)} \text{ satisfies } \varphi(\mathrm{x}/t)],$$

and in this way a truth-condition for a mixed quantificational sentence like $(\forall x)(\Sigma x)\varphi$ is ultimately determined.

Kripke notes that the valid sentences of the pure predicate calculus (PC) with objectual quantifiers are precisely the valid sentences when quantifiers are construed substitutionally. There is a perfect parallel between objectual validity and substitutional validity. The language of PC can be notationally adapted for either interpretation. It has denumerably many variables; write them italicized for the objectual treatment, roman for the substitutional. It has denumerably many n-adic predicate letters and n-adic function letters, for each n. (As usual, 0-adic predicate letters are the sentence letters of propositional logic; 0-adic function letters are, in effect, "derived" constants. The language of PC is contains no "constants" properly so-called; its only constants are these 0-adic function letters.) Also present are parentheses, connectives & and $\sim$, and one of $\exists$ and Σ. We thus have two languages, perspicuous notational variants of each other, but the same point could be made by means of a single language and the two semantical apparatuses applied to it.

First the objectual case. A sentence φ of PC is *objectually valid* iff it is true on any objectual interpretation, i.e.,

(i) for any nonempty universe U of values of variables,

(ii) for any assignment of members of U to φ's constants,

(iii) for any assignment of sets of n-tuples of members of U to φ's n-adic predicate letters,

(iv) for any assignment of n-ary functions $U^n \rightarrow U$ to φ's n-adic function letters.

(Since the only constants are the 0-ary function letters, clause (*ii*) is a special case of clause (*iv*); it is included for explicitness.)

Now the substitutional case. An n-adic *functor* of a language is an expression with at most n number of free variables, such that any substitution of terms from C for the free variables results in a term. The 0-ary functors are the constants. A sentence φ of the pure predicate calculus is *substitutionally valid* iff it is true for any substitutional extension L' of any L,

(i) for any (nonempty) substitution-class C,

(ii) for any (uniform) replacements of φ's constants with any members of C,

(iii) for any (uniform) replacements of φ's n-adic predicate letters with any n-adic atomic formulas of L',

(iv) for any (uniform) replacements of φ's n-adic function letters with any n-adic functors of L.

This perfect parallelism shows why the set of objectually valid sentences of PC is precisely the set of substitutionally valid ones. Kripke notes (336):

> The important fact is this: one and the same formal system can be used to prove all the valid formulae of both the pure referential and the pure substitutional predicate calculi. If a formula is valid in the one sense, it is valid in the other.

The result does not conflict with our earlier observation on the relative strengths of objectual quantifiers and their substitutional counterparts, however. It does not follow that for a language containing both sorts of quantifiers, objectual quantifications and their substitutional counterparts are equivalent. Their relationships are as indicated earlier, due to the “problem of too few names”.

Kripke intended his observations to answer objections lodged against substitutional quantification in previous influential work (Tharp 1971; Wallace 1971, 1972, and elsewhere). These had argued that that substitutional quantification was both formally and meaning-theoretically defective. It is generally accepted that Kripke showed such quantification to be formally unproblematic on those particular grounds, but subsequent theorists have continued to object to it on further meaning-theoretical grounds. We turn now to a family of these.

THE LYCAN-VAN INWAGEN-FINE OBJECTION(S)

William L. Lycan [1979] raises a serious meaning-theoretical worry concerning substitutional quantification. A similar concern is expressed by Peter van Inwagen [1985], who gives a different argument for almost the same conclusion. Lycan concludes that substitutional quantifications don’t *say* anything at all; van Inwagen holds that *if* substitutional quantifications do not say anything *metalinguistic*, then he does not know what they say (nor does anyone else), so that indeed they say nothing.

Lycan’s argument runs in general outline as follows. Recall Frege’s horizontal operator. Applied to a sentence, the horizontal yields an expression denoting the True if the sentence itself denotes the True (ie., is true), and denotes the False if the sentence itself denotes the False (is false). Applied to any other expression (which denotes something other than a truth-value if it denotes anything), the resulting expression also denotes the False. Thus

(10) $——7$

denotes the false. In this sense, then, (10) has a truth-condition. Note, however, that (10) does not *say* anything, in a very natural sense of “say”. Now compare (10) with (11), where the substitution-class for the quantifier contains only the left-hand parenthesis.

(11) $(\Sigma x)x\varphi \rightarrow \varphi)$

This is not gibberish. Its sole substitution-instance is $(\varphi \rightarrow \varphi)$. Since this is true, (11) is true. But again, (11) fails to *say* anything. By the same token, then,

(12) $(\Sigma x)(x \text{ is a dog})$

fails to say anything, despite its truth. To reach this conclusion, Lycan first considers (12) against the background of a semantics whose substitution-class *C* of terms available as substituends for (12)'s quantifier is *motley*: it includes terms such that their substitutions for x in "x is a dog" are syntactically acceptable, but also some that are not.

A semantics with a motley substitution-class is not obviously unworkable. The idea is that the members of C which do not yield well-formed wffs upon substitution are simply irrelevant semantically, for the truth-condition of a substitutional existential quantification is an objectual existential quantification over C, rendering syntactically inappropriate members of C semantically immaterial. Lycan urges the intuition that (12) is no more able to say anything than is (11). Then, given this, he concludes that even when C contains only terms that uniformly yield grammatical instances, (12) remains "semantically mute," as he puts it. The transition from the motley case to the uniform case is made by means of the consideration that if the motley C contains, say, both names and the left-hand parenthesis, then both (12) and

(13) (Σx)xBetsy is a dog)

are thus made true by the same substitution-instance,

Betsy is a dog

and it is difficult to see how (12) can avoid muteness in light of (13)'s.

This argument is intuitively powerful. It seems clear enough that (11) expresses no proposition, despite its possession of a truth-condition. Having a truth-condition is no guarantee of an expression's possession of content. Lycan distinguishes the two possible treatments of (12) because he expects this intuition to be shared concerning (12) when interpreted against a motley C, and then argues that there is at best a formal and meaning-theoretically *arbitrary* difference between this treatment and a treatment of (12) where C is uniform. Lycan holds that an arbitrary, formal change in the semantics such as restricting C to grammatically acceptable substituends for x cannot bestow meaning upon (12). Thus (12) is mute on the uniform substitutional treatment as well as the motley one.

Van Inwagen [1985] complains that no one knows what proposition a sentence like (12) expresses (unless it is the metalinguistic objectual quantification over terms, which it is supposed not to express). Substitutional quantifications are radically unlike their corresponding objectual quantifications, for which it is easy to give the "bottom-level" expressed propositions that are not *about* language and truth.

Van Inwagen's conclusion is based on the *assumption* that (12) does not *mean* that "x is a dog" has a true instance, while Lycan's conclusion is that (12) doesn't mean anything at all, even *that.* Van Inwagen concludes only that neither he nor anyone else understands (knows which proposition is expressed by) (12).

A few words about what a sentence *means*, or rather what it needn't mean, are in order. A correct T-sentence for a given sentence φ is a metalinguistic biconditional whose left-hand side attributes truth to φ and whose right-hand side states the condition that must obtain in order for φ to be true. When φ is a bottom-level sentence, an adequate truth theory must provide a T-sentence for φ whose right-hand is also a bottom-level sentence of the semantical metalanguage. Many theorists use something like this idea, "truth-conditionalists" in particular, who cannot avail themselves of Tarski's requirement that the right-hand side be a *translation* of φ into the metalanguage. This would *presuppose* a notion of meaning, and they take their project to be, roughly, a characterization of meaning in terms of the antecedently better-understood notion of truth. Because they cannot appeal to the *meaning* of φ in placing stronger constraints on the right-hand sides of T-sentences, there are correct T-sentences which have no chance of expressing on their right-hand sides what the sentence in question *means.* If (14) is correct, then so is (15).

(14) $\mathrm{Tr}(\varphi)$ iff there is a dog.

(15) $\mathrm{Tr}(\varphi)$ iff [there is a dog, and $(\psi \vee \sim \psi)$]

Any truth theory that issues in (14) also issues in (15) (assuming its logic is classical, a matter we will not pursue). But proposition-theorists insist that propositions, which are supposed to be the meanings of sentences, are individuated more finely than by logical equivalence. The proposition that there is a dog is not the proposition that-there-is-a-dog-and-either-ψ-or-not-ψ (where ψ is some unrelated sentence). Thus although (14) and (15) associate φ with the same truth-condition, they do not both identify the *proposition* that φ expresses. The mere fact that the truth-condition for a substitutional existential quantification is given by means of metalinguistic objectual quantification is no reason to think that the proposition it expresses is a metalinguistic one. Van Inwagen grants that the truth-condition of (12) can be given metalinguistically, and requests more information: what *bottom-level* proposition does (12) express, whereby a metalinguistic statement of its truth-condition can be seen to be correct?

Now, van Inwagen's argument presupposes that if (12) does express a proposition, then we should be able to say what (12) means. He grants that among the correct T-sentences for (12) should be one whose right-hand side expresses the very proposition that (12) expresses. Assuming (12) doesn't *mean* that there is a term that yields a truth when substituted for x in "x is a dog," then *we cannot say* what (12) says. Call the standard statement of (12)'s truth-condition, the objectual quantification over terms, "∃". Van Inwagen writes [p. 285],

> [(12)], therefore, does not express ∃. But then what proposition does it express? I do not know. No one knows. No one knows because no one has ever *said.* I know of only one proposition — ∃ — that one might reasonably suppose...to be the proposition [(12)] expresses, and I am told by those who know what proposition [(12)] expresses if *anyone* knows what proposition [(12)] expresses that ∃ is *not* the proposition [(12)] expresses...That is, I do not understand [(12)]. And neither, I think, does anyone else. But then neither I nor anyone else understands substitutional quantification.

Lycan, unlike van Inwagen, is not happy with the use of propositions in an account of what a sentence says. He is a truth-conditionalist who thinks that truth-conditions are at least a good starting point for an account of understanding a sentence (knowing what it means). Knowing what a sentence means is not just a matter of knowing a T-sentence for it, but of being able to compute a T-sentence on the basis of the sentence's structure and its words, where it is assumed that one knows the meaning of its words and how its compositional structure contributes to truth-conditions. Substitutional quantification threatens the whole truth-conditionalist project, since anyone familiar with Σ and Π knows how to compute a T-sentence for (12) in Lycan's preferred way.

Michael Hand [1993] suggests a response to these arguments. First, there is a consideration that comes close to answering van Inwagen without answering Lycan. Van Inwagen seeks a bottom-level proposition expressed by (12), and finds one for its objectual counterpart by means that fail for (12). He explains our grasp of $\forall$ and $\exists$ by reference to the quantificational idioms of natural language. He takes it that the proposition expressed by

(16) $(\exists x)(x \text{ is a dog})$

is also expressed in ordinary English by "There is a dog," so that his answer to the question what is meant by (16) is simple. (16) means that there is a dog. This indicates that the root of his problem is simply that English (and any other natural language he understands) lacks substitutional quantifiers. If English already had its own vernacular substitutional existential quantifier, say "sigma" (to be used in constructions like "There is *sigma* F" in contrast to "There is *an* F"), or if we add "sigma" to English, yielding a new, augmented "natural language," then it would be easy enough to produce a natural language sentence expressing the same proposition that (12) expresses. Speaking this augmented language, we can answer van Inwagen's demand to be told what proposition is expressed: (12) says that there is sigma dog. To resist this option, van Inwagen must deny that any proposition is expressed (in the new language) by this new sentence. Yet it is not a difficult imaginative exercise to become fluent enough in the new enlargement of English to think twice about this denial. It amounts to a claim that the expansion of English is not a possible natural language.

A response to Lycan is also in the offing, which further undermines van Inwagen's predicted denial. Assuming that the semantics of this new locution is best formulated by means of a uniform substitution-class, and that speakers have become fluent with their new substitutional quantifier, their grasp of natural language substitutional quantification appears to be on a meaning-theoretical par with their grasp of English objectual quantification. A Lycanian truth-conditionalist can therefore grant that (12) is no more mute than (16). That is, Lycan's considerations on muteness do not count against a sophisticated truth-conditionalist theory of meaning. (Hand tries to finesse Lycan's worry about motley *vs.* uniform substitution-classes.)

In much the same vein, Kit Fine [1989] argues that our *understanding* of substitutional quantifications is to be explained in terms of our *understanding* of objectual quantification. He considers $(\Sigma \mathrm{x})F\mathrm{x}$ and its corresponding objectual quantification $(\exists x)Fx$, interpreting F in the latter in a new, "shifted" way as the predicate holding of a term t just when Ft, interpreted in the original way, is true. The substitutional quantifier is therefore read as a metalinguistic one over terms: its *meaning* is objectual, though metalinguistic. Like van Inwagen, Fine reports difficulty in understanding $(\Sigma \mathrm{x})F\mathrm{x}$ in any other way than by means of the "shifted" interpretation of F, and concludes that there is apparently no difference in our understanding $(\Sigma \mathrm{x})$ and our understanding of $(\exists x)$: rather the difference between $(\Sigma \mathrm{x})F\mathrm{x}$ and $(\exists x)Fx$ is in our interpretation of the predicate F.

Fine's reinterpretation of $(\exists x)Fx$ is of course just the metalinguistic characterization of the truth-condition of $(\Sigma \mathrm{x})F\mathrm{x}$ in new guise, and his argument is no stronger than van Inwagen's own, though Fine remains uncommitted to a proposition-theoretic account of meaning. His argument is cast in terms of our *understandings* of these sentences, and thus is neutral on the divide between proposition-theoretic meaning theories and sophisticated truth-conditionalist ones. It nonetheless appears subject to a rephrasing of Hand's response to van Inwagen's case: Fine cannot understand $(\Sigma \mathrm{x})F\mathrm{x}$ except by means of his metalinguistic "shift" because he does not understand the corresponding new vernacular statement "There is sigma F", which by hypothesis does not engage an interpretive "shift" with respect to the predicate.

THE PROBLEM OF TOO MANY NAMES

H. A. Lewis [1985] observes that certain *nonstandard* substitutional quantifiers give rise to an problem reminiscent of the problem of too few names. It is the problem of too *many* names. As a special case, it plagues numerical substitutional quantifiers. Consider the nonstandard quantifier "most" and a straightforward substitutional account, assuming a straightforward referential semantics. (We write nonstandard objectual quantifiers as well as their variables in italics, and nonstandard substitutional ones along with their variables in roman. We reserve "term" for names belonging to C.) Where φ has at most x free,

$$(\mathrm{most}\ \mathrm{x})\varphi \text{ is true iff } (\underline{\mathit{most}}\ t \in \mathrm{C})(\varphi(\mathrm{x}/t) \text{ is true})$$

An objectual interpretation of "most" is predictable; here we do not require that φ has at most x free (for parallelism with the clauses for $\forall$ and $\exists$).

$$\boldsymbol{s} \text{ satisfies } (\mathit{most}\ x)\varphi \text{ iff } (\underline{\mathit{most}}\ \alpha \in \mathrm{U})(\boldsymbol{s}^{(x/\alpha)} \text{ satisfies } \varphi)$$

Let our universe contain only three members: $\mathrm{U}=\{a,b,c\}$. Let C, however, contain ten terms, $\{t_1,...t_{10}\}$, with t_1 naming a, t_2 naming b, and the remainder naming c. Let F's extension be $\{a, b\}$. Then the objectual quantification

$$(\mathit{most}\ x)Fx$$

is clearly true, for two out of the universe's three members belong to F's extension. Yet among the ten substitution-instances of Fx are a single true one for a, a single true one for b, and eight that are false in virtue of c's absence from F's extension. That is, the substitutional quantification

(most x)Fx

is false. The example is easily modified so that the substitutional quantification is spuriously true — let a and b each have ten names, and c only one, but only c belong to F's extension.

Many nonstandard quantifiers give rise to this problem, including "several", "a few", and "many". The problem also afflicts numerical quantifiers like "exactly three", as is easily seen. (As remarked earlier, identity must be treated with kid gloves in the presence of substitutional quantification, but as Kripke observed, under certain assumptions it behaves. The key condition is transparency, so we stipulate that the language has no referential opacity.)

Lewis concludes that the use of substitutional quantification requires a semantical principle limiting the language to a single name for each member of U. This would solve his problem, but it is an extreme solution, since on this account any identity statement having distinct names flanking the identity sign is false. Such extremity is avoidable when the base semantics of the language is referential, as it clearly is in Lewis's scenario.

The problem shows that substitutional quantification cannot be relied upon to "count" individuals correctly. Rather, these nonstandard quantifiers count *terms*. The problem arises because a single individual may be counted more than some other individual because it has more terms denoting it, and its many terms yield substitution-instances that are all true or false together.

Let δ be the denotation function taking each term to its value, the member of U it denotes. For each $t \in$C, there is the equivalence class $\{t' \in C : \delta(t') = \delta(t)\}$ of terms that are coreferential with t. Given a well-ordering of terms, each equivalence class has a least member. Call these the *proper terms* of C, and consider the set $C*$ of C's proper terms. $C*$ has no more than one term per member of U (though it may leave some out entirely — the problem of too *few* names). In this way we ensure that these substitutional quantifiers do not count individuals more than once. Instead of requiring that every true identity-statement of the language has the same name on both sides, we ensure, in effect, only that no true identity statement has distinct *proper* terms on both sides. This less stringent solution tweaks the semantics of the substitutional quantifiers, rather than tweaking the language itself.

(Σx)φ is true iff ($\underline{\exists} t \in$C*)($\varphi$(x/$t$) is true)

(most x)φ is true iff ($\underline{\textit{most}}$ $t \in$C*)(φ(x/t) is true)

and so on.

Both Lewis's proposal and this one exploit the referentiality of the (base) semantics. Both proposals involve sensitizing the truth-conditions of substitutional nonstandard quantifiers to features of the denotation function δ, the first by requiring δ to be one-to-one, and the latter by reducing C to C*, thereby requiring δ to be one-to-one over proper terms. The need for such a move counts as evidence that the natural language nonstandard quantifiers themselves ought always to be construed, regardless of the discourse in which they are used, as objectual. When they are used in a discourse in which the standard quantifiers "there are", "some," "every," etc. are also used, this would constitute a good reason to treat these as objectual, too. In particular, it counts against the substitutional strategy for avoiding ontological commitment. For instance, any case for construing the ordinary quantifiers as substitutional in discourse about fictitious entities like Pegasus is weakened by the observation that nonstandard quantifiers are also used to talk about fictitious entities: "Leopold Bloom was taller than most of his countrymen."

NATURAL LANGUAGE SUBSTITUTIONALISM

We turn finally and briefly to a legendary view of which it is difficult to find advocates, the view that the ordinary quantifiers "there is" and "for all" (and their variants) of natural language are themselves substitutional in nature. The view has an idealistic air, and appears to bear on the claim that we understand the use of $\forall$ and $\exists$ by means of our grasp of ordinary language quantifiers, which are supposedly objectual. Can the objectuality of ordinary language quantifiers be called into question *generally*? Kripke (380) observes that the question of whether ordinary language quantifiers are objectual or substitutional evaporates.

> [A]nyone who doubts or denies that English has any resources for making genuinely existential assertions has cut the ground out from under his feet; he has no way even of *formulating* the question at issue. The alleged "issue", then is bogus.

Indeed, what is this substitutionalism to make of the obvious claim that the truth of "There is a cat in my kitchen" does not require that any cats in my kitchen have names?

BIBLIOGRAPHY

[Benacerraf, 1983] P. Benacerraf. Mathematical truth. In Paul Benacerraf and Hilary Putnam (eds.), *Philosophy of Mathematics: Selected Readings*, 2nd edition. Cambridge: Cambridge University Press, 1983.

[Dummett, 1973] M. Dummett. *Frege: Philosophy of Language*. Cambridge, Mass.: Harvard University Press, 1973.

[Fine, 1989] K. Fine The problem of De Re modality. In Joseph Almog, John Perry, Howard Wettstein (eds.), *Themes From Kaplan*. New York: Oxford University Press, 1989.

[Gottlieb, 1980] D. Gottlieb *Ontological Economy: Substitutional Quantification and Mathematics*. Oxford: Clarendon Press, 1980.

[Hand, 1993] M. Hand. "Truth, meaning, and substitutional quantification. *Philosophical Studies* 68, 195–216, 1993.

[van Inwagen, 1981] P. van Inwagen. Why I don't understand substitutional quantification. *Philosophical Studies* 39, 281–85, 1981.

[Kripke, 1976] S. Kripke. Is there a problem about substitutional quantification? In Gareth Evans and John McDowell (eds.), *Truth and Meaning*. Oxford: Clarendon Press, 1976.

[Lewis, 1985] H. A. Lewis. Substitutional quantification and nonstandard quantifiers. *Noûs* 19, 447–451, 1985.

[Lycan, 1979] W. G. Lycan. Semantic competence and funny functors. *The Monist* 62, 209–22, 1979.

[Marcus, 1993] R. B. Marcus. Modalities and intensional languages, in her *Modalities*. New York: Oxford University Press, 1993.

[Parsons, 1983] C. Parsons. A plea for substitutional quantification, in his *Mathematics in Philosophy*. Ithaca: Cornell University Press, 1983.

[Quine, 1996] W. V. O. Quine. Three grades of modal involvement, in his *The Ways of Paradox and Other Essays*. New York: Random House, 1966.

[Quine, 1980] W. V. O. Quine. On what there is, in his *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press, second edition revised, 1980.

[Quine, 1969] W. V. O. Quine. Existence and quantification, in his *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969.

[Tennant, 1979] N. Tennant. *Antirealism and Logic*. Oxford: Clarendon Press, 1979.

[Tharp, 1971] L. Tharp. Truth, quantification and abstract objects. *Noûs* V, 363–79, 1971.

[Wallace, 1972] J. Wallace. On the frame of reference. In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*. Dordrecht: Reidel, 1972.

[Wallace, 1971] J. Wallace. Convention T and substitutional quantification, *Nous* V, 199–212, 1971.

MANY-VALUED LOGICS

Siegfried Gottwald

1 BASIC IDEAS

1.1 From classical to many-valued logic

Logical systems in general are based on some formalized language which includes a notion of well-formed formula, and then are determined either semantically or syntactically.

That a logical system is semantically determined means that one has a notion of *interpretation* or *model*[1] in the sense that w.r.t. each such interpretation every well-formed formula has some *(truth) value* or represents a function into the set of (truth) values. It furthermore means that one has a notion of validity for well-formed formulas and, based upon it, also a natural *entailment relation* between sets of well-formed formulas and single formulas (or sometimes also whole sets of formulas).

That a logical system is syntactically determined means that one has a notion of *proof* and of provable formula, i.e. of (formal) *theorem*, as well as a notion of *derivation from* a set of *premisses.*

From a philosophical, especially epistemological point of view the semantic aspect of (classical) logic is more basic than the syntactic one, because it are mainly the semantic ideas which determine what are suitable syntactic versions of the corresponding (system of) logic.

The most basic (semantic) assumptions of classical, i.e. two-valued – propositional as well as first-order — logic are the principles of bivalence and of compositionality. Here the *principle of bivalence* is the assumption that each sentence[2] is either true or false under any one of the interpretations, i.e. has exactly one of the truth values $\top$ and $\bot$, usually numerically coded by 1 and 0. And the *principle of compositionality*[3] is the assumption that the value of each compound well-formed formula is a function of the values of its (immediate) subformulas.

The most essential consequence of the principle of compositionality is the fact that each one of the propositional connectives as well as each one of the (first

---

[1] We prefer to use the word interpretation in general and to restrict the use of the word model to particular interpretations which are tied in a specific way with sets of well-formed formulas.

[2] By a sentence one means either any well-formed formula of the corresponding formalized propositional language, or any well-formed formula of the corresponding formalized first order language which does not contain any free individual variable.

[3] Sometimes this principle is also named principle of extensionality.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

order) quantifiers is semantically determined by a function (of suitable arity) from the set of (truth) values into itself, or by a function from its powerset into the set themselves.

Disregarding the quantifiers for a moment, i.e. restricting the considerations to the propositional case, the most essential (semantical) point is the determination of the truth value functions, i.e. of the operations in the truth value set which characterize the connectives. From an algebraic point of view, hence, the crucial point is to consider not only the *set* of truth values, but a whole *algebraic structure* with the truth value set as its support. And having in mind that all the classical connectives are definable from the connectives for conjunction, disjunction, and negation, this means to consider the set $\{0, 1\}$ of truth values together with the truth functions $\min, \max$ and $1 - \dots$ of these connectives — and this is a particular Boolean algebra. The semantical notion of validity of a well-formed formula φ w.r.t. some interpretation now means that φ has truth value 1 at this particular interpretation. (And universal validity of course means being valid for every interpretation.)

Generalizing the notions of truth value, of interpretation and of validity in such a way that the truth value structure may be any (nontrivial) Boolean algebra $\mathcal{B} = \langle B, \sqcap, \sqcup, ^*, \mathbf{0}, \mathbf{1} \rangle$, that an interpretation is any mapping from the set of all propositional variables into B, the truth value functions for conjunction, disjunction, negation are chosen as $\sqcap, \sqcup, ^*$, respectively, and validity of a well-formed formula φ w.r.t. a given interpretation means that φ has the Boolean value $\mathbf{1}$ at this interpretation.

This generalized type of interpretations can easily become extended to the first order case: one then has to consider only complete Boolean algebras and has to consider the operations of taking the infimum or supremum as the operations corresponding to the universal or existential quantifier.

It is well-known that the class of universally valid formulas of classical logic is just the class of all formulas valid for all (nontrivial) — and complete (in the first order case) — Boolean algebras as truth value structures. This fact is referred to by saying that the class of (complete) Boolean algebras is characteristic for classical logic.

Many-valued logic deviates from the two basic principles of bivalence and of compositionality only in that it neglects the principle of bivalence. Therefore, any system S of many-valued logic is characterized (i) by a suitable *formalized language* $\mathcal{L}_\mathsf{S}$ which comprises

- its (nonempty) family $\mathcal{J}^\mathsf{S}$ of (basic) *propositional connectives*,
- its (possibly empty) family of *truth degree constants*,
- its set of *quantifiers*,[4]

---

[4] Each quantifier for simplicity is supposed here to be a unary one. This means that we allow to have some kinds of generalized quantifiers in the sense of Mostowski [Mostowski, 1957] but we do not consider the possibility to have quantifiers with more than one scope, as is allowed e.g. in [Rescher, 1969; Gottwald, 2001].

and adopts the usual way of defining the class of well-formed formulas w.r.t. these syntactic primitives, and parallel to these syntactic data (ii) by the corresponding semantic data, i.e. by

- a (nonempty) set $\mathcal{W}^{\mathsf{S}}$ of *truth degrees*,[5]
- a family of *truth degree functions* together with a correspondence between these truth degree functions and the propositional connectives of the (formal) language,
- a (possibly empty) family of nullary operations, i.e. of elements of the truth degree set together with a one-one correspondence between the members of this family and the truth degree constants of the (formal) language,
- a set of *quantifier interpreting functions* from the power set $I\!P(W^{\mathsf{S}})$ of $\mathcal{W}^{\mathsf{S}}$ into $\mathcal{W}^{\mathsf{S}}$ together with a one-one correspondence between these functions and the quantifiers of the (formal) language.

Based upon these data which constitute any particular system S of many-valued logic, it is a routine matter to combine with each truth degree evaluation e of all atomic formulas an extension $\mathsf{Val}^{\mathsf{S}}$ which gives for each well-formed formula φ of the language of S its truth degree $\mathsf{Val}^{\mathsf{S}}(\varphi, e)$ under e.

As usual, $\mathsf{Val}^{\mathsf{S}}(\varphi, e)$ could be defined by recursion on the complexity of φ.

And, again as usual, in the propositional case such an evaluation e is given immediately, and in the first-order case $e = (\mathfrak{A}, v)$ is determined by a suitable interpretation $\mathfrak{A}$ together with an assignment v of objects from $\mathfrak{A}$ to the (individual) variables of the language.

1.2 Particular truth degree sets

Usually one additionally assumes that the classical truth values (or some "isomorphic" copies of them, also coded by 1 and 0) appear among the truth degrees of any suitable system S of many-valued logic:

(1) $\{0, 1\} \subseteq \mathcal{W}^{\mathsf{S}}$.

Formally, for the systems S of many-valued logic there is essentially no restriction concerning the set $\mathcal{W}^{\mathsf{S}}$ of truth degrees of S besides (1). Nevertheless the choice of $\mathcal{W}^{\mathsf{S}}$ as a set of numbers (either integers or rationals or even reals) is widely accepted use. At least as long as one is not interested to have an ordering of the truth degrees which allows for incomparable truth degrees[6]. The existence of incomparable truth degrees, however, may be crucial for certain particular applications, e.g. in a situation where the truth degrees are intended to code parallel evaluations of different points of view. To imagine such a situation assume to be interested, in image processing, to evaluate whether a certain point P belongs to a certain figure $\mathcal{F}$, i.e. to determine the truth value of the sentence "P is a point of $\mathcal{F}$". For black and white pictures, the evaluation yields one of the truth values $\top, \bot$. Being instead confronted with a graytone picture, the evaluation of this sentence may yield as truth degrees the values of some scale which characterizes the different gray levels. And being, finally, confronted with a colored picture, e.g. on the screen of some monitor, which consists of (colored) pixels which themselves are generated by superposing pixels of the three basic colors, then it may be reasonable to evaluate the above mentioned sentence by a truth degree which is a triple of the levels of intensity of the basic colors which give point P.

---

[5] For systems of many-valued logic we prefer to call their semantic values truth *degrees* to emphasize the difference to the truth values of classical logic. Moreover this term appears to be a bit more neutral concerning ontological commitments. And this is important because one does not have any preferred ontological reading for the truth degrees of many-valued logics. The intuitive understanding of these degrees completely depends upon the particular applications under consideration. Because this is a completely different situation compared with classical logic, this change in the terminology is intended to underline just this difference.

Another widely accepted kind of approach is to assume that among the truth degrees there is a smallest one, usually interpreted as an equivalent for the truth value $\bot$, and a biggest one, usually interpreted as an equivalent for $\top$.

Based on these common assumptions it is usually at most a simple matter of isomorphic exchange of the structure of truth degrees to assume (1) together with

$$(2) \quad \mathcal{W}^{\mathsf{S}} \subseteq [0,1] \subseteq \mathbb{R}.$$

And this choice of the truth degree set shall be the standard one in the following discussions.

For the case of infinitely many truth degrees it is common usage to consider either countably many or uncountably many truth degrees and furthermore to choose either one of the truth degree sets

$$(3) \quad \mathcal{W}_0 =_{def} \{x \in \mathbb{Q} \,|\, 0 \leq x \leq 1\} \quad \text{or} \quad \mathcal{W}_\infty =_{\text{def}} \{x \in \mathbb{R} \,|\, 0 \leq x \leq 1\}.$$

For the case of finite sets of truth degrees usually one additionally assumes that these truth degrees form a set of equidistant points of the real unit interval $[0,1]$, i.e. are of the kind

$$(4) \quad \mathcal{W}_m =_{\text{def}} \{\frac{k}{m-1} \mid 0 \leqq k \leqq m-1\}$$

for some integer $m \geqq 2$.

1.3 Designated truth degrees

In classical logic there is a kind of superiority of the truth value $\top$ over the other one $\bot$: given some well-formed formula (or some set of formulas) one mainly is interested in those interpretations for the system of many-valued logic one is working in which make the given formula(s) true.

---

[6] Of course, even in such a situation one can take a set of numbers for the set of truth degrees — and adjoin another ordering relation to these numbers than their natural ordering. But this is rather unusual.

With the set $\mathcal{W}^{\mathsf{S}}$ of truth degrees each system S of many-valued logic has its equivalent to the truth value set $\{\top, \bot\}$. However, even admitting condition (1), this does not mean that the truth degree 1 has to be *the* equivalent of $\top$. The rather vague philosophical idea that in many-valued logic one considers some kind of "splitting" of the classical truth values does not by itself determine which truth degrees "correspond" to $\top$. Therefore, to determine some system S of many-valued logic does not only mean to fix its set of truth degrees and its formal language, i.e. its connectives, quantifiers, predicate symbols and individual constants, together with their semantic interpretations, it also means to fix which truth degrees "correspond" to $\top$.

Formally this means, that with each system S of many-valued one connects not only its set $\mathcal{W}^{\mathsf{S}}$ of truth degrees but also some set $\mathcal{D}^{\mathsf{S}}$ of *designated* truth degrees. Of course, usually one supposes

$$(5) \quad 1 \in \mathcal{D}^{\mathsf{S}} \subseteq \mathcal{W}^{\mathsf{S}} \quad \text{together with} \quad 0 \notin \mathcal{D}^{\mathsf{S}}.$$

Such a choice of designated truth degrees is of fundamental importance for the generalization of the notions of logical validity and of logical consequence.

1.4 Logical validity and logical consequence

Based on the notion of designated truth degrees it is essentially a routine matter to generalize the notions of logical validity and of logical consequence to any system S of many-valued logic.

Adopting the standard usage to mean by a *sentence* either a well-formed formula in the case of a propositional language $\mathcal{L}_{\mathsf{S}}$ or a well-formed formula without free individual variables of a first order language $\mathcal{L}_{\mathsf{S}}$, one calls a sentence φ *valid* w.r.t. an evaluation e of the atomic formulas iff it has a designated truth degree for that evaluation, and one calls a sentence φ *logically valid* iff it is valid for every suitable evaluation.

As usual, furthermore, a *formula* φ of a first order language $\mathcal{L}_{\mathsf{S}}$ is called *valid in a* given *interpretation* $\mathfrak{A}$, iff it has a designated truth degree w.r.t. any assignment of objects of $\mathfrak{A}$ to the individual variables of the language $\mathcal{L}_{\mathsf{S}}$.

On the other hand, an evaluation e is called a *model* of a formula φ iff this formula φ is valid under this evaluation e. And e is called a *model* of a set Σ of well-formed formulas iff it is a model of any formula $H \in \Sigma$. The fact that e is a model of φ, or of Σ, as usual is denoted by $e \models \varphi$ or $e \models \Sigma$. And, as usual, one considers the (crisp) *model classes*

$$(6) \quad \mathsf{Mod}^{\mathsf{S}}(\Sigma) =_{\text{def}} \{e \mid e \models \Sigma\} = \{e \mid \mathsf{Val}^{\mathsf{S}}(\varphi, e) \in \mathcal{D}^{\mathsf{S}} \text{ for all } \varphi \in \Sigma\}.$$

As usually, hence, one has that the logical validities are just those formulas φ whose model class is the class of all evaluations.

Sometimes, the notion of model is generalized a bit further in many-valued logic. Given some truth degree α and some formula φ, an evaluation e is called an *α-model* of φ iff the truth degree of φ under the evaluation e equals α or[7] iff it is greater or equal to α. We prefer here, to speak in the last mentioned case of a $(\geq \alpha)$-model.

Correspondingly the notion of α-model of a set of sentences is used. In this case, the most suitable way is to call an evaluation an $(\geq \alpha)$-model of a set Σ of sentences iff this interpretation is an $(\geq \alpha)$-model of each sentence $\varphi \in \Sigma$.

Based on these preliminaries, the notion of logical consequence is defined almost in the standard way, but again with mainly two slightly different basic intuitions. For simplicity, we restrict again the considerations to sentences only. The extension to all well-formed formulas happens in many-valued logic exactly as in classical logic.

One defines that a sentence φ is a *logical consequence* of a set Σ of sentences, usually also written $\Sigma \models \varphi$, iff

(*Version 1*): each model of Σ is also a model of φ, or equivalently

$$(7) \quad \Sigma \models \varphi \iff \mathsf{Mod}^{\mathsf{S}}(\Sigma) \subseteq \mathsf{Mod}^{\mathsf{S}}(\varphi)\,.$$

(*Version 2*): each $(\geq \alpha)$-model of Σ is also a $(\geq \alpha)$-model of φ.

With $\models \varphi$ as shorthand for $\emptyset \models \varphi$ in any case $\models \varphi$ means that φ is logically valid.

The checking of well-formed formulas of any propositional system of many-valued logic for being logically valid can be done in the same way as for classical logic by determining complete truth degree tables — and this is effective provided the set of truth degrees is finite.

THEOREM 1. *For each finitely many-valued system* S *of propositional logic the property of being a logically valid sentence is decidable, and for each finite set* Σ *of sentences of* S *also the property of being a logical consequence of* Σ *is decidable.*

2 OUTLINE OF THE HISTORY

Many-valued logic as a separate part of logic was created by the works of J. Łukasiewicz [1920] and E.L. Post [1921] in the beginning 1920th. Admittedly, both authors have not been the first ones which did not assume the principle of bivalence, but earlier attempts to do logic without this principle of bivalence did not prove to be influential[8].

---

[7] Both these variants are in use. One has to check the use of the term α-model in the particular case to see which version applies.

[8] Even the previous paper [Łukasiewicz, 1913] of J. Łukasiewicz which also admitted generalized truth "degrees" besides the traditional truth values $\top, \bot$ did not influence the development of logic toward many-valued logic in any perceivable manner.

The prehistory of many-valued logic, however, may be traced back up to Aristotle[9] who e.g. in his *De Interpretatione*, chap. 9, discussed the problem of future contingencies, i.e. the problem which truth value a proposition should have today which asserts some future event. This problem was stimulating even for J. Łukasiewicz [1930], and it is closely tied with the philosophical problem of determinism. The link is provided by the interpretation that the classification of some future event as (actually) "possible" or "undetermined" may well be seen as the acceptance of a third "truth value" besides $\top$ and $\bot$. Surely, this reading is not the necessary one. Nevertheless, the ancient philosophical school of Epicureans which tended toward indeterminism refused the principle of bivalence, whereas the school of the Stoics did accept it — and strongly advocated determinism.

The same problem of *contingentia futura* was also the source for several extended discussions during the Middle Ages, cf. e.g. [Łukasiewicz, 1935; Michalski, 1937; Baudry, 1950; Rescher, 1969], without getting resolved. And in the phase of the general revival of investigations into the field of logic during the second half of the 19th century the idea of neglecting the principle of bivalence appeared (partly without clear mentioning of this fact) to H. McColl [McColl, 1897], cf. also [Lovett, 1900], and Ch.S. Peirce, cf. [Peirce, 1931, vol. 4] as well as [Fisch and Turquette, 1966; Turquette, 1967].

The real starting phase of many-valued logic was the time interval from about 1920 till about 1930, and the main force of development was the Polish school of logic under J. Łukasiewicz. The papers [Łukasiewicz and Tarski, 1930; Łukasiewicz, 1930] as well as the influential textbook [Lewis and Langford, 1932], all published in 1930, explain the core ideas as well as the background of philosophical ideas and the main technical results proven up to this time.[10] They also stimulated further research into the topic.

In [Łukasiewicz, 1930] Łukasiewicz intends to give a modal reading to his many-valued propositional logic, claiming that only the 3-valued and the infinite valued case (with the set of all rationals between 0 and 1 as truth degree set) are really of interest for applications. In [Łukasiewicz and Tarski, 1930] however all finitely many-valued propositional systems and the just mentioned infinitely many-valued one are discussed, always based on a negation and an implication connective as primitive ones characterized semantically by their truth degree functions.

Parallel with this "Polish" development the US-American mathematician E.L. Post [Post, 1921] designed a family of finitely valued systems. However, it seems that his main aim was not a philosophical one, but "only" a technical: he was interested in the problem of functional completeness, i.e. in the representation of *all* truth degree functions (say from $\mathcal{W}_m$ into $\mathcal{W}_m$) with the help of only a few of them.

---

[9] The interested reader may consult e.g. [Łukasiewicz, 1930; Łukasiewicz, 1970; Rescher, 1969; Patzig, 1973].

[10] As a side remark it has to be mentioned that P. Bernays [Bernays, 1926] used more than the usual two truth values of classical logic to study independence problems for systems of axioms for systems of classical propositional calculus. But in his case these multiple values were only formal tools for his unprovability results.

Basic theoretical results for systems of many-valued logics which followed the initial phase of "Polish" many-valued logic have e.g. been

- M. Wajsberg's [1977] axiomatization of the three valued (propositional) system L_3 of Łukasiewicz, i.e. of that one propositional system with Łukasiewicz's implication and negation connectives as primitive connectives,
- the extension of Łukasiewicz's system L_3 to a functionally complete one and its axiomatization by J. Słupecki [1936],
- the work of K. Gödel [Gödel, 1932] and S. Jaśkowski [Jaśkowski, 1936] which clarified the mutual relations of intuitionistic and many-valued logic in the sense that it was proven that there does not exist a single (propositional) many-valued system whose set of logically valid formulas coincides with the set of logically valid formulas of intuitionistic (propositional) logic,
- the application of systems of three valued logic to the problems of logical antinomies by Bočvar [1938; 1943] with the third truth value read as "senseless",
- the application of systems of three valued logic to problems of partially defined function by S. Kleene [1938; 1952] with the third truth value read as "undefined".

Furthermore, during the 1940th basic approaches have been generalized and essential results were proven by J. B. Rosser and A.R. Turquette in a series of papers and later on most of this material collected in their monograph [Rosser and Turquette, 1952], which besides the Łukasiewicz papers of 1930 was the standard reference for years. At least up to 1969 in which year the nice monograph [Rescher, 1969] appeared which became one of the standard references to many-valued logics up to the 1990s.

In the 1950s and 1960s then there was some decline in the interest in many-valued logics — and at the same time some shift in the focus of what were considered as the more interesting problems. The decline as well as the shift may have been caused by the same situation: the fact — mentioned in the Rosser/Turquette monograph [Rosser and Turquette, 1952] quite open — that up to this time no really convincing applications had been approached by the methods of many-valued logic. The shift in interest inside many-valued logic thus happened toward problems of definability of operations in $\{1, 2, \ldots, n\}$ from particular sets of such operations, i.e. toward problems connected with the functional incompleteness of most of the then "usual" systems of connectives of (finitely) many-valued systems of (propositional) logic. In the background, however, there was not only the theoretical problem of functional completeness, there was (and is) also the related problem in switching theory of sets of suitable elementary circuits which allow to generate all the (finitary) operations in $\{1, 2, \ldots, n\}$ by being combined into suitable circuits — or the related problem to determine the class of all operations which can be generated by some given ones.

3 BASIC SYSTEMS OF MANY-VALUED LOGICS

If one looks systematically for many-valued logics which have been designed for quite different applications, one finds four main types of systems:

- the Łukasiewicz logics L_κ as explained in [Łukasiewicz, 1930];
- the Gödel logics G_κ from [Gödel, 1932];
- the product logic Π studied in [Hájek *et al.*, 1996];
- the Post logics P_m for $2 \leq m \in \mathbb{N}$ from [Post, 1921].

The first two types of many-valued logics each offer a uniformly defined family of systems which differ in their sets of truth degrees and comprise finitely valued logics for each one of the truth degree sets (4) together with an infinite valued system with truth degree set (3), which formally is indicated by choosing $\kappa \in \{n \in \mathbb{N} \mid n \geq 2\} \cup \{\infty\}$. For the fourth type an infinite valued version is lacking.

In their original presentations, these logics look rather different, regarding their propositional parts. For the first order extensions, however, there is a unique strategy: one adds a universal and an existential quantifier such that quantified formulas get, respectively, as their truth degrees the infimum and the supremum of all the particular cases in the range of the quantifiers.

As a reference for these and also other many-valued logics in general, the reader may consult [Gottwald, 2001].

3.1 The Gödel logics

The simplest ones of these logics are the *Gödel logics* G_κ which have a conjunction $\wedge$ and a disjunction $\vee$ defined by the minimum and the maximum, respectively, of the truth degrees of the constituents:

$$(8) \quad u \wedge v = \min\{u, v\}\,, \qquad u \vee v = \max\{u, v\}\,.$$

For simplicity we denote here and later on the connectives and the corresponding truth degree functions by the same symbol.

These Gödel logics have also a negation $\sim$ and an implication $\rightarrow_{\mathsf{G}}$ defined by the truth degree functions

$$(9) \quad \sim u = \begin{cases} 1\,, & \text{if } u = 0; \\ 0\,, & \text{if } u > 0. \end{cases} \qquad u \rightarrow_{\mathsf{G}} v = \begin{cases} 1\,, & \text{if } u \leq v; \\ v\,, & \text{if } u > v. \end{cases}$$

The systems differ in their truth degree sets: for each $2 \leq \kappa \leq \infty$ the truth degree set of G_κ is $\mathcal{W}_\kappa$.

3.2 The Łukasiewicz logics

The *Łukasiewicz logics* L_κ, again with $2 \le \kappa \le \infty$, have originally been designed in [Łukasiewicz, 1930] with only two primitive connectives, an implication $\rightarrow_{\mathsf{L}}$ and a negation $\neg$ characterized by the truth degree functions

$$(10)\ \ \neg u = 1 - u\ , \qquad u \rightarrow_{\mathsf{L}} v = \min\{1, 1 - u + v\}\,.$$

The systems differ in their truth degree sets: for each $2 \le \kappa \le \infty$ the truth degree set of G_κ is $\mathcal{W}_\kappa$.

However, it is possible to define further connectives from these primitive ones. With

$$(11)\ \ \varphi \,\&\, \psi =_{\mathrm{df}} \neg(\varphi \rightarrow_{\mathsf{L}} \neg\psi)\ , \qquad \varphi \veebar \psi =_{\mathrm{df}} \neg\varphi \rightarrow_{\mathsf{L}} \psi$$

one gets a (strong) conjunction and a (strong) disjunction with truth degree functions

$$(12)\ \ u \,\&\, v = \max\{u + v - 1, 0\}\,, \qquad u \veebar v = \min\{u + v, 1\}\,,$$

usually called the *Łukasiewicz* (arithmetical) *conjunction* and the *Łukasiewicz* (arithmetical) *disjunction*. It should be mentioned that these connectives are linked together via a De Morgan law using the standard negation of this system:

$$(13)\ \ \neg(u \,\&\, v) = \neg u \veebar \neg v\,.$$

With the additional definitions

$$(14)\ \ \varphi \wedge \psi =_{\mathrm{df}} \varphi \,\&\, (\varphi \rightarrow_{\mathsf{L}} \psi) \qquad \varphi \vee \psi =_{\mathrm{df}} (\varphi \rightarrow_{\mathsf{L}} \psi) \rightarrow_{\mathsf{L}} \psi$$

one gets another (weak) conjunction $\wedge$ with truth degree function min, and a further (weak) disjunction $\vee$ with max as truth degree function, i.e. one has the conjunction and the disjunction of the Gödel logics also available.

3.3 The Product logic

The *product logic* Π, in detail explained in [Hájek *et al.*, 1996], has a fundamental conjunction $\odot$ with the ordinary product of reals as its truth degree function, as well as an implication $\rightarrow_\Pi$ with truth degree function

$$(15)\ \ u \rightarrow_\Pi v = \begin{cases} 1\,, & \text{if } u \le v; \\ \frac{u}{v}\,, & \text{if } u < v. \end{cases}$$

Additionally it has a truth degree constant $\overline{0}$ to denote the truth degree zero.

In this context, a negation and a further conjunction are defined as

$$(16)\ \ \sim \varphi =_{\mathrm{df}} \varphi \rightarrow_\Pi \overline{0}\ , \qquad \varphi \wedge \psi =_{\mathrm{df}} \varphi \odot (\varphi \rightarrow_\Pi \psi)\,.$$

Routine calculations show that both connectives coincide with the corresponding ones of the infinite valued Gödel logic G_∞. And also the disjunction $\vee$ of this Gödel logic becomes available, now via the definition

(17) $\varphi \vee \psi =_{\mathrm{df}} ((\varphi \rightarrow_{\Pi} \psi) \rightarrow_{\Pi} \psi) \wedge ((\psi \rightarrow_{\Pi} \varphi) \rightarrow_{\Pi} \varphi)\,.$

There is, however, no natural way to combine with this (infinite valued) product logic a whole family of finite valued systems by simply restricting the set of truth degrees to some $\mathcal{W}_m$ as in the previous two cases: besides $\mathcal{W}_2$ no such set is closed under the ordinary product, and for $\mathcal{W}_2$ the product coincides e.g. with the minimum operation.

3.4 The Post logics

The Post system P_m for $m \geq 2$ has truth degree set $\mathcal{W}_m$. These propositional systems have been originally formulated uniformly in negation and disjunction as basic connectives with the following truth degree functions:

$$\sim u = \begin{cases} 1\,, & \text{for } u = 0\,, \\ u - \frac{1}{m-1}\,, & \text{for } u \neq 0\,, \end{cases} \qquad u \vee v = \max\{u, v\}\,.$$

Contrary to the previous systems, the definition of negation here does not seem to be given in a uniform way independent of the number of truth degrees. However, it is always just a cyclic permutation of all the truth degrees (in their natural order).

For the sets of designated truth degrees a canonical choice does not exist; already Post [Post, 1921] has discussed the possibility that there may be chosen truth degrees different from 1 as designated ones. Nevertheless, $\mathcal{D}^P = \{1\}$ is a kind of standard choice.

The set of basic connectives of each one of the Post systems P_m is functionally complete, i.e. allows to represent every possible truth degree function (over $\mathcal{W}_m$). Therefore each one of the Post systems P_m, with $\mathcal{D}^P = \{1\}$ as the set of designated truth degrees, covers its corresponding Łukasiewicz system with the same set of truth degrees — in the sense that the set of L_m-tautologies is a subset of the set of P_m-tautologies, and that this set of P_m-tautologies does not contain any formula φ whose Łukasiewicz negation $\neg H$ is L_m-satisfiable, of course always via a suitable reading of the Łukasiewicz connectives in the Post systems. And the same holds true for the corresponding m-valued Gödel system G_m.

If one enriches all the finitely many-valued (propositional) Łukasiewicz systems L_m with truth degree constants for all their truth degrees, then these enriched systems L_m^* become functionally complete. And this means that the extended m-valued Łukasiewicz systems L_m^* and the m-valued Post logics become interdefinable (for each fixed number m of truth degrees). Hence there is in principle no essential difference between both types of (finitely valued) systems: all what can be expressed in the "Post world" can also be expressed in the (extended) "Łukasiewicz world", and vice versa.

4 STANDARD AND ALGEBRAIC SEMANTICS

The fundamental many-valued logics have their *standard semantics* as explained: the sets $\mathcal{W}_m$ or the whole real unit interval $[0, 1]$ as truth degree sets, and the connectives (and quantifiers) as mentioned.

And derived from these basic choices one has the notions of validity under some evaluation, of logical validity, and of model as explained.

Besides these standard semantics, all these many-valued logics have also *algebraic semantics* determined by suitable classes $\mathcal{K}$ of truth degree structures. The situation is similar here to the case of classical logic: the logically valid formulas in classical logic are also just all those formulas which are valid in all Boolean algebras.

Of course, these structures have to have the same signature as the language $\mathcal{L}$ of the corresponding logic. This means that these structures provide for each connective of the language $\mathcal{L}$ an operation of the same arity, and they have to have — in the case that one discusses the corresponding first order logics — suprema and infima for all those subsets which may appear as value sets of formulas. Particularly, hence, they have to be (partially) ordered, or at least pre-ordered.

For each formula φ of the language $\mathcal{L}$ of the corresponding logic, for each such (generalized truth degree) structure $\mathbf{A}$, and for each evaluation e which maps the set of atomic formulas of $\mathcal{L}$ into the carrier of $\mathbf{A}$, one has to define a value $\mathsf{Val}(\varphi, e)$, and finally one has to define what it means that such a formula φ is *valid in* $\mathbf{A}$. Then a formula φ is *logically valid* w.r.t. this class $\mathcal{K}$ iff φ is valid in all structures from $\mathcal{K}$.

The standard way to arrive at such classes of structures is to start from the Lindenbaum algebra of the corresponding logic, i.e. its algebra of formulas modulo the congruence relation of logical equivalence. For this Lindenbaum algebra one then has to determine a class of similar algebraic structures which — ideally — forms a variety.

A *variety* is a class $\mathcal{K}$ of algebraic structures which is equationally definable, i.e. for which there exists a set $\mathcal{E}$ of equations between terms of the language of these structures such that an algebraic structure $\mathbf{A}$ belongs to $\mathcal{K}$ iff $\mathbf{A}$ is a model of $\mathcal{E}$. Besides this characterization in logical terms there is also a characterization in purely algebraic terms: a variety is a class $\mathcal{K}$ of algebraic structures which is closed under the formations of subalgebras, of homomorphic images, and of direct products. For the algebraic details the interested reader may e.g. consult [Burris and Sankappanavar, 1981; Denecke and Wismath, 2002; Grätzer, 1979].

4.1 Gödel and Łukasiewicz logics

It is remarkable that for both these types of many-valued logics corresponding algebraic semantics have mainly been developed for the infinite valued systems, and have been considered in the context of completeness proofs.

For the infinite valued Gödel logic G_∞ such a class of structures is, according to the completeness proof given by Dummett [Dummett, 1959], the class of all *Heyting algebras*, i.e. of all relatively pseudo-complemented lattices, which satisfy the pre-linearity condition

$$(18)\quad (u \rightarrowtail v) \sqcup (v \rightarrowtail u) = \mathbf{1}\,.$$

Here $\sqcup$ is the lattice join and $\rightarrowtail$ the relative pseudo-complement.

For the infinite valued Łukasiewicz logic L_∞ the corresponding class of structures is the class of all *MV-algebras*, first introduced again within a completeness proof by Chang [Chang, 1958], and more recently extensively studied in [Cignoli *et al.*, 2000].

And for the product logic the authors of [Hájek *et al.*, 1996] introduce a class of lattice ordered semigroups which they call *product algebras*.

It is interesting to recognize that all these structures — prelinear Heyting algebras, MV-algebras, and product algebras — are abelian lattice ordered semigroups with an additional "residuation" operation.

For the Łukasiewicz as well as for the Gödel infinite valued logics these algebraic semantics have long been considered as a mathematically nice, but logically not really important tool.

The recent development of infinite valued logics, which shall be discussed later on, beginning with §6, has completely modified this point of view: this development got very important stimulations from suitable algebraic semantics.

For the finite valued logics from both families, separately developed algebraic semantics did not yet find considerable interest. There was, again in the context of a completeness proof, an approach by R.S. Grigolia [Grigolia, 1977] toward MV-algebras for m-valued Łukasiewicz logics, called MV$_m$-algebras.[11] But these structures play only a marginal rôle in algebraic investigations toward many-valued logics.

Besides a certain minor interest in this topic, this situation may mainly be caused by the fact that the finite valued Łukasiewicz logics L_m become functionally complete after enriching their language either with truth degree constants for all truth degrees, or with suitable unary (and binary) connectives which are sufficient to characterize each one of the truth degrees. And for these enriched systems G.C. Moisil [Moisil, 1940; Moisil, 1941], and later on R. Cignoli [Cignoli, 1982], offered algebraic semantics. These *Łukasiewicz algebras*, however, are quite difficult structures which do not have a primitive counterpart for the Łukasiewicz implication. And therefore there is a tendency to consider them more as variants of Post algebras, as explained in §4.3, than as "natural" counterparts to MV-algebras.

---

[11] It seems, to the best of this authors knowledge, that for the finite valued Gödel logics there are no separate algebraic studies. In principle, however, Heyting-algebras with m elements should do the job for G_m.

4.2 Product logic

The product logic, as introduced in [Hájek *et al.*, 1996], was from the very beginning designed as a logic which had, in parallel, a standard semantics — provided by the real unit interval and by a product based conjunction as a fundamental connective — as well as an algebraic semantics, formed by the class of all product algebras — introduced in [Hájek *et al.*, 1996] again within a completeness proof.

We shall not explain more details here because this whole approach proved to become paradigmatic for the development of t-norm based infinite valued logics, a topic which shall be discussed later on, starting with §6.

4.3 Post logics

Contrary to the situation for the Łukasiewicz and the Gödel systems, for the Post systems in their original form there exist only very few syntactically oriented studies toward constituting or investigating logical calculi for these systems. Instead, for the Post systems one mainly was interested in corresponding algebraic structures, which were suitable to form an algebraic semantics, and investigated such structures earlier, and in more detail, as similar structures for the Łukasiewicz and the Gödel systems. Rosenbloom in a paper [Rosenbloom, 1942] of 1942 was the first one to do this. His algebraic structures shall here be called P-algebras for short — but not be considered in detail: the interested reader may e.g. consult [Gottwald, 2001].

One of the main reasons for the difficulty and complexity of the defining conditions of P-algebras is the fact that the Post systems as well as the P-algebras have only two primitive notions, their connectives resp. their basic operations, but have maximal expressive power in the sense of being functionally complete. That this choice of the primitive notions really is the main obstacle toward a simplification became clear as Epstein [Epstein, 1960] in 1960 changed these basic operations and found a much simpler class of "essentially" these P-algebras, now called Post algebras. The reservation "essentially" here comes from the fact that formally the choice of other basic operations creates another type of algebraic structures, and it means–in more technical terms– that Post algebras and P-algebras are definitionally equivalent. We shall not go into details and refer e.g. again to [Gottwald, 2001] for details.

What are not covered by these basic considerations are possible infinite valued generalizations of these logical calculi, or of these Post algebras. Approaches toward this problem started e.g. with papers on generalizations of the notion of Post algebras like [Cat-Ho, 1973; Dwinger, 1968; Dwinger, 1977; Traczyk, 1967]. The most influential paper, however, which also discussed the corresponding logical systems was the paper [Rasiowa, 1973] of Rasiowa in which Post algebras of order $\omega + 1$ and corresponding systems of infinitely many-valued (first-order) logic have been introduced. The algebraic theory of these Post algebras of order $\omega + 1$ is partly given in [Rasiowa, 1973].

Another such infinitely many-valued generalization of the standard Post systems is discussed e.g. in [Epstein and Rasiowa, 1990; Epstein and Rasiowa, 1991], Post algebras of order $\omega + \omega^*$.

The Post algebras of finite or infinite order and the systems of many-valued logic related with them seem to be of particular importance for investigations in computer science, which rely on many-valued logic as a toolbox, because these Post systems are functionally complete and well suited to study the representability of truth degree functions on the basis of some predetermined set of basic truth degree functions, as determined e.g. by available electronic components, cf. [Rine, 1977] for a good introduction.

5 PARTICULAR THREE- AND FOUR-VALUED SYSTEMS

Each system of many-valued logic which is not only intended to be some particular kind of formalism, but supposed to express some meaning, is confronted with the problem to offer a meaning for its truth degrees, or at least for its "additional" truth degrees different from the degrees 1 for "true" and 0 for "false". In the light of this problem, three-valued as well as four-valued systems get particular importance because for them only one or two "additional" truth degrees exist and need an interpretation. Hence it should not appear as a surprise that from the viewpoint of philosophically-oriented applications three- and four-valued systems assumed a more prominent role than other systems, at least as other finitely many-valued systems.

5.1 Three-Valued Systems

Here we shall be interested only in some such 3-valued systems and mainly restrict the attention to two strongly related systems introduced by Bočvar [1938] and Kleene [1938], and a system designed as the "true" logic of the natural language by Blau [1978]. The intentions of these authors, connected with their systems, have been quite different despite some strong similarities of the systems.

The main problem of Bočvar has been the philosophical and logical analysis of logical and semantical antinomies as they appear in first-order and higher-order logic, often in connection with some lack of care e.g. in the use of the (set theoretic) comprehension principle or of metatheoretical notions, cf. e.g. [Beth, 1965]. Therefore his preferred interpretation of the additional truth degree $\frac{1}{2}$ was its reading as "meaningless", "paradoxical", or "senseless". The starting point of Kleene, on the other hand, was a mathematical one and related to his research on partial recursive relations. Such relations sometimes may be undefined. Therefore in his case the intended reading of the additional truth degree $\frac{1}{2}$ was "undefined" or "undetermined". Both systems coincide in their approach to consider the truth degrees $1, 0$ just as the counterparts of the classical truth-values $\top, \bot$. Accordingly for both systems the degree 1 is the only designated one.

Bočvar subdivides his truth degree functions and thus also his connectives into *internal* and *external* ones. The characteristic property of the internal truth degree functions is that they have always a (truly: the) function value different from $0,1$ if a argument value differs from $0,1$. The external truth degree functions, on the other hand, map into $\{0,1\}$. Hence this subdivision is not a (complete) classification because there obviously exist truth degree functions which are neither external nor internal ones. The the 3-valued system $\boldsymbol{B}_3$ of Bočvar has four basic connectives $\neg$, $\wedge_+$, J_0, J_1 for internal negation, internal conjunction, external negation, and external assertion. The internal negation $\neg$ is nothing but the negation of the Łukasiewicz system L_3. The internal conjunction may, again in terms of L_3, be defined as

(19) $\varphi \wedge_+ \psi =_{\text{def}} (\varphi \wedge \psi) \vee (\varphi \wedge \neg\varphi) \vee (\psi \wedge \neg\psi)$

or by the formula

$$u \wedge_+ v =_{\text{def}} \begin{cases} \min(u,v)\,, & \text{if } u,v \in \{0,1\}\,, \\ \frac{1}{2}\,, & \text{if } u = \frac{1}{2} \text{ or } v = \frac{1}{2}\,. \end{cases}$$

The truth degree functions of J_0, J_1 are given by

$$\mathsf{J}_0(u) = \begin{cases} 1\,, & \text{if } u = 0 \\ 0\,, & \text{otherwise} \end{cases} \qquad \mathsf{J}_1(u) = \begin{cases} 1\,, & \text{if } u = 1 \\ 0\,, & \text{otherwise} \end{cases}$$

Thus $\boldsymbol{B}_3$ is a subsystem of L_3. It is a proper subsystem because it was shown in [Finn, 1974] that the Łukasiewicz implication $\to_{\mathsf{L}}$ is not $\boldsymbol{B}_3$-definable.

Further internal versions of disjunction, implication and biimplication can be defined as

$$\begin{aligned} \varphi \vee_+ \psi &=_{\text{def}} \neg(\neg\varphi \wedge_+ \neg\psi)\,, && (20) \\ \varphi \to_+ \psi &=_{\text{def}} (\varphi \wedge_+ \neg\psi), && (21) \\ \varphi \leftrightarrow_+ \psi &=_{\text{def}} (\varphi \to_+ \psi) \wedge_+ (\psi \to_+ \varphi)\,, && (22) \end{aligned}$$

and get truth degree functions which coincide over $\{0,1\}$ with their classical counterparts and are internal truth degree functions. Corresponding external versions of these connectives result by a uniform approach which e.g. for external conjunction and external disjunction reads as

$$\begin{aligned} \varphi \Cap \psi &=_{\text{def}} \mathsf{J}_1(\varphi) \wedge_+ \mathsf{J}_1(\psi)\,, \\ \varphi \Cup \psi &=_{\text{def}} \mathsf{J}_1(\varphi) \vee_+ \mathsf{J}_1(\psi)\,. \end{aligned}$$

The axiomatizability problem for $\boldsymbol{B}_3$ has been discussed and solved in [Finn, 1971; Finn, 1974a]. We shall not treat it here.

The 3-valued system $\boldsymbol{K}_3$ of Kleene has the so-called *strong* connectives

(23) $\neg, \wedge, \vee, \to_{\text{K}}, \leftrightarrow_{\text{K}}$.

The first three of them are again the (equally denoted) connectives of L_3. The last two of them are characterized by the truth degree tables:

$\to_{\mathrm{K}}$	0	$\frac{1}{2}$	1
0	1	1	1
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
1	0	$\frac{1}{2}$	1

$\leftrightarrow_{\mathrm{K}}$	0	$\frac{1}{2}$	1
0	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
1	0	$\frac{1}{2}$	1

Also for this system $\boldsymbol{K}_3$ the connectives $\neg, \wedge, \vee$ are just the same as in the Łukasiewicz system L_3.

This system $\boldsymbol{K}_3$ furthermore has *weak* connectives $\wedge_+, \vee_+, \to_+$ which coincide with the equally denoted connectives of the Bočvar system $\boldsymbol{B}_3$, i.e. which yield formulas which have truth degree $\frac{1}{2}$ iff one of their constituents has truth degree $\frac{1}{2}$. These weak connectives are definable from the connectives of the list (23) because of (19), (20), and (21).

Because the implication $\to_{\mathrm{K}}$ of the Kleene system is L_3-definable, e.g. by

$$(24) \quad \varphi \to_{\mathrm{K}} \psi =_{\mathrm{def}} (\varphi \to_{\mathsf{L}} \psi) \wedge (\varphi \vee \neg\varphi \vee \psi \vee \neg\psi)\,,$$

also the 3-valued Kleene system $\boldsymbol{K}_3$ is a subsystem of the Łukasiewicz system L_3. Again, however, the Łukasiewicz implication $\to_{\mathsf{L}}$ is not definable in the Kleene system and this system therefore is a proper subsystem of L_3. This undefinability follows from the fact that each (binary) connective $\vartriangle$ which is definable from the basic connectives in the list (23) has a corresponding truth degree function which assumes the value $\frac{1}{2}$ if all its arguments have this value, and hence cannot be the truth degree function of the Łukasiewicz implication.

If one combines the basic connectives of the Bočvar system with the basic connectives of the Kleene system, then $\to_{\mathsf{L}}$ becomes definable as mentioned by Šestakov [Šestakov, 1964]. Therefore neither are all the connectives of the Kleene system definable in the Bočvar system, nor conversely are all the connectives of the Bočvar system definable in the Kleene system.

Systems of many-valued propositional logic with an intended reading of the third truth degree as "meaningless", or as "undefined", or as something like, as in the Bočvar and Kleene systems, and with basic connectives which essentially can be defined in L_3, have also been investigated e.g. in [Åqvist, 1962; Ebbinghaus, 1969; Halldén, 1949; Piróg-Rzepecka, 1977; Segerberg, 1965]. In [Bolc and Borowik, 1992] a longer, but concise survey of such and other 3-valued systems is given. Additionally the interested reader should also consult [Goddard and Routley, 1973].

It is interesting to notice that the intended reading of the third truth degree as "undefined" in the Kleene system, and the connection with the similar approach via truth value gaps which we mentioned earlier, was a substantial fact that this system was more recently considered in connection with partial, i.e. sometimes undefined, truth predicates e.g. in [Gupta and Belnap, 1993; Kripke, 1975; Martin and Woodruff, 1975; McGee, 1991].

Another approach toward 3-valued systems comes from the consideration of vague predicates like "hot water", i.e. of predicates which in some cases neither really apply nor really do not apply to some objects. This effect can be modeled in different ways. A usual one is via fuzzy sets, which supported the recent investigations into infinite valued systems, as explained later on. But also 3-valued systems provide a (rough) possibility — with the third truth degree read as "neither completely applies nor fully does not apply". From this point of view three-valued logic was studied with a philosophical attitude quite early in [Black, 1937; Black, 1963] and more recently e.g. in [Körner, 1966] and also in [Blau, 1978; Cleave, 1974; Kearns, 1974; Kearns, 1979].

It should additionally be mentioned that the phenomenon of presuppositions was discussed within the realm of truth value gaps, as well as in the realm of 3-valued — but also of 4-valued — systems. And also systems of paraconsistent logic have been discussed which can be based on finitely many truth degrees, cf. e.g. [Epstein, 1990; Sette, 1973; Sette and Carnielli, 1995].

Both of these aspects, i.e. vague predicates as well as presuppositions, are covered in the use of three-valued logic for the analysis of natural language given by Blau [1978]. His approach is essentially based on discussions in the realm of the philosophy of language. From these discussions he gets the fundamental motivations for the intuitive understanding of the three truth degrees. He identifies the classical truth value $\top$ with the truth degree 1, and he splits the classical truth value $\bot$ into the two degrees 0 and $\frac{1}{2}$. He takes the truth degree 0 as a modified version of the (usual) truth value $\bot$, and his intended reading for the truth degree $u = \frac{1}{2}$ is "undetermined", combined with the understanding that the appearance of this degree is caused either by the use of vague predicates, or by the use of nondenoting names, i.e. by reference to unsatisfied presuppositions.

On the propositional level his considerations are based on three primitive connectives, two kinds of negation connectives $\neg, \approx$ and a conjunction connective $\wedge$. Again $\neg$ is the Łukasiewicz negation of L_3, and $\wedge$ the weak conjunction of that system. The additional negation $\approx$ is characterized by the truth degree function

$$(25)\quad \approx\!(u) = \begin{cases} 0\,, & \text{if } u = 1\,, \\ 1\,, & \text{if } u \neq 1\,. \end{cases}$$

All these connectives satisfy the normal condition, hence this system is not functionally complete. However, these connectives suffice to introduce some further, interesting connectives, e.g. the Gödel negation $\sim$ as

$$\sim \varphi =_{\mathrm{def}} \neg\!\approx\!\neg\varphi\,,$$

and to introduce an implication connective $\rightarrow_{\mathsf{Bl}}$ by the formula

$$\varphi \rightarrow_{\mathsf{Bl}} \psi =_{\mathrm{def}} \approx\!\varphi \vee \psi\,.$$

The main background idea behind the choice of this implication connective is the authors claim that only this implication connective is suitable for a three-valued modeling of (two-valued) sentences of the form "All A are B" in natural language, of course read as short form for: all objects which have property A

also have property B. And this claim is essentially based upon the idea that for the truth of a sentence of this form "all objects which have property A also have property B" it is completely out of any rational interest to allow the antecedent "a particular object has property A" to be undetermined. Formally this means that inside a universally quantified sentence the case that the antecedent has truth degree $\frac{1}{2}$ should not be a reason that the whole sentence may become not true.

A core intuitive point of the whole approach is the idea that natural language uses only such connectives which satisfy the normal condition, because the appearance of the truth degree $u = \frac{1}{2}$ for some sentenceφ(in some particular situation) is simply an unintended mistake caused either by the use of vague predicates, or by the use of nondenoting names.

5.2 Four-Valued Systems

In contrast to the situation with three-valued systems, where there are a lot of approaches and interpretations, only a few approaches concern four-valued systems and give particular interpretations to the four truth degrees. One of these rare exceptions is Łukasiewicz [1953] who, in his later years, preferred a four-valued approach via his system L_4 toward a modal reading of the truth degrees over his original three-valued one via L_3 in [Łukasiewicz, 1920].

However, instead of this reference to the system L_4 with its linearly ordered truth degree set $\mathcal{W}_4$, an approach toward four-valued systems has become prominent more recently which makes essential use of the natural partial ordering of the truth degree set $\mathcal{W}_4^* = \{0,1\}^2 = \{(0,0),(0,1),(1,0),(1,1)\}$, and which connects a very natural interpretation with these degrees.

This approach was inspired by theoretical work on systems of relevance logic and later on also applied to considerations on how to treat — possibly inconsistent — information in computers, e.g. in data bases or knowledge bases, like in [Belnap, 1977; Belnap, 1977a; Dunn, 1976; Dunn, 1998], but also to discussions related to the liar paradox, cf. [Visser, 1984].

In data and knowledge bases information is stored, e.g. in the form of "facts", i.e. sentences which are marked as true or false. This information may have been collected from different sources, and at different times. One kind of use which can be made from such information is that one asks such data or knowledge bases, or the computers they are stored in, to answer some suitable questions, e.g. for (the confirmation or refutation of) simple statements concerning facts. The crucial point is that this information usually is incomplete — and often even inconsistent. Therefore one should allow a computer to answer not only "true" or "false", but also "I don't know" — and even "true and false". Of course, the answer "I don't know" indicates incompleteness of the information stored in the data or knowledge base, and the answer "true and false" indicates that this information is inconsistent — in the simplest case because contradictory facts have been stored.

For the computer these answers are just "marks" which it has to connect with sentences (which e.g. formulate the questions it is asked to answer). And with these four “values” the computer should also be able to “reason” internally, because one likes (within a bit more sophisticated applications) that the computer is not only able to repeat something that he was told before, he should also be able to connect different facts by a kind of (internal) reasoning mechanism.

Of course, as this explanation shows, this is an epistemic understanding of these four “values”, and not an ontological one: the “real world” is, of course, treated here as well covered by the basic ideas of two-valued logic. This remark does not say, however, anything about the suitability of a four-valued approach based on this (epistemic) understanding of the truth degrees.

It is interesting to notice that some (preliminary) form of an ontological understanding of these four degrees appears in ancient Indian logic, e.g. in the work of Sanjaya who worked prior to the sixth century B.C. There the principle of bivalence was rejected in favor of an idea of so-called “four corners”, which correspond just to an ontological reading of these four degrees, cf. [Horn, 1989; Raju, 1954].

For simpler reading we write $\mathbf{T} = (1, 0)$, $\mathbf{F} = (0, 1)$, $\mathbf{N} = (0, 0)$, $\mathbf{B} = (1, 1)$, and understand the truth degree $\mathbf{N}$ (for “none”) as indicating “underdetermination” or a *gap*, i.e. the lack of information on a truth value, and the truth degree $\mathbf{B}$ (for “both”) as “overdetermination” or a *glut*, i.e. the presence of contradictory information on some truth value.

One has even more than only the truth degree set $\mathcal{W}_4^*$ in this case: one has a natural (partial) ordering of these degrees, having in mind that they evaluate the (computers) knowledge about the truth value of some sentence φ. Then surely the degree $\mathbf{T} = (1, 0)$ is ranked in top position, and the degree $\mathbf{F} = (0, 1)$ is ranked in lowest position — because it is most preferable to have a (definitely) true sentence φ, and worse to have a (definitely) false one. The two other degrees $\mathbf{N} = (0, 0)$, $\mathbf{B} = (1, 1)$ are ranked somehow “between” the degrees $\mathbf{T}, \mathbf{F}$ because they, in some suitable sense, if assigned to φ, leave open both possibilities that φ may “really” be true or be false. As a result, this provides the truth degree set $\mathcal{W}_4^*$ with a lattice structure as indicated in Fig. 1.
In this truth degree lattice $\mathcal{W}_4^*$ the lattice ordering $\leqq$ goes “bottom-up”, i.e. lattice elements which are on a lower level position are smaller ones.

With the corresponding lattice operations $\sqcap, \sqcup$ one has natural candidates for truth degree functions for a conjunction and a disjunction connective $\curlywedge, \curlyvee$ of a (propositional) system D_4 of four-valued logic which is to be based on the intuitions discussed up to now. And this choice fits even well into the intuitive picture mentioned previously.

This intuitive picture provides also the basis for the introduction of a negation connective $\rightharpoondown$ into D_4, with a similar relationship to classical negation, as $\curlywedge$ and $\curlyvee$ have to classical conjunction and disjunction, respectively. This essentially means that to a formula $\rightharpoondown H$ the truth value $\bot$ (or: $\top$) should be assigned if to φ the truth value $\top$ (or: $\bot$) is assigned. So one immediately has for $\rightharpoondown$ the truth degree table given in Fig. 2.

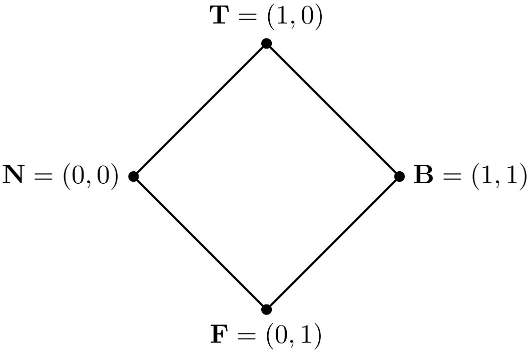


Figure 1. Four-valued truth degree lattice $\mathcal{W}_4^*$

φ	**F**	**N**	**B**	**T**
$\neg H$	**T**	**N**	**B**	**F**

Figure 2. Truth table characterization of the D_4-negation

These connectives $\curlywedge, \curlyvee, \neg$ form the basic vocabulary of D_4.

One can also prove that the two three-valued subsystems of D_4, which are constituted by the restrictions of the truth degrees to $\{\mathbf{T}, \mathbf{F}, \mathbf{N}\}$ or $\{\mathbf{T}, \mathbf{F}, \mathbf{B}\}$, coincide as three-valued systems, and are subsystems of L_3.

The crucial point now is to define a suitable entailment relation $\models_{\mathsf{D}}$ which fits well into this intuitive realm.

A very natural approach seems to be, to say again that a set of formulas Σ entails a formula φ iff each model of Σ is also a model of φ. However, what shall we understand by a model of a set Σ of formulas? Well, nothing but some valuation which gives to all the formulas of Σ a designated truth degree.

So the problem arises what the designated truth degrees should be. Up to now there is no agreement on this point. One of the possible approaches is to take only $\mathbf{T}$ as a designated truth degree, i.e. to put $\mathcal{D}^{\mathsf{D}} = \{\mathbf{T}\}$.

In this case the resulting notion of model fits well into the background intuition: for our computer a model of a set of formulas should be any (partial and non-functional) two-valued valuation which makes all the formulas of Σ definitely true — with "definitely true" understood here as meaning *true but not false or valueless.* And this represents the standard notion of a model. Thus we get in this case according to (7):

$$(26) \quad \Sigma \models_{\mathsf{D}} H \ \Leftrightarrow\ \mathsf{Mod}^{\mathsf{D}}(\Sigma) \subseteq \mathsf{Mod}^{\mathsf{D}}(H)\,,$$

with the notion of model class defined as in (6).

With applications to relevance logic in mind, Dunn [Dunn, 1976] considers instead both truth degrees $\mathbf{T}, \mathbf{B}$ as designated. The intuition behind this choice is that a formula which has such a designated truth degree is considered as "at least true".[12] Problems of definability of truth degree functions, i.e. of connectives, and of relations between truth degrees for this choice of designated truth degrees have been discussed e.g. in [Avron, 1999; Pynko, 1999].

The models in the sense of [Dunn, 1976] for this situation are just the ($\geqq \mathbf{B}$)-models, with $\leqq$ for the lattice ordering of the truth degree lattice $\mathcal{W}_4^*$. Let us call them *weak models* for the moment, and denote the class of all weak models of some set Σ of formulas by $\mathsf{Mod}_0^{\mathsf{D}}(\Sigma)$.

Then Dunn discusses the following notion $\models_{\mathsf{D}}^0$ of *weak entailment*:

$$\Sigma \models_{\mathsf{D}}^0 H \;\Leftrightarrow\; \mathsf{Mod}_0^{\mathsf{D}}(\Sigma) \subseteq \mathsf{Mod}_0^{\mathsf{D}}(H)\,,$$

which is more general than the previous notion from (26), as the following result shows.

PROPOSITION 2. *For each set Σ of formulas of D_4 and each formula φ one has*

$$\Sigma \models_{\mathsf{D}}^0 H \quad \Rightarrow \quad \Sigma \models_{\mathsf{D}} H\,.$$

Still another notion of entailment $\models_{\mathsf{D}}^*$ is taken into account in [Belnap, 1977; Belnap, 1977a]. There Belnap considers however only entailment relationships of the form $\varphi \models_{\mathsf{D}}^* H$. This corresponds for the two former cases to a restriction to finite sets Σ of formulas. The definition of $\models_{\mathsf{D}}^*$ is the following:

$$\varphi \models_{\mathsf{D}}^* \psi \quad =_{\mathrm{def}} \quad \mathsf{Val}^{\mathsf{D}}(\varphi, \beta) \leqq \mathsf{Val}^{\mathsf{D}}(\psi, \beta) \text{ for all valuations } \beta.$$

It is immediately clear that one has for all formulas φ, ψ of D_4:

$$\varphi \models_{\mathsf{D}}^* \psi \quad \Rightarrow \quad \varphi \models_{\mathsf{D}}^0 \psi\,.$$

For this notion of entailment Belnap gives in [Belnap, 1977; Belnap, 1977a] a finite list of principles claiming that it is a complete list to infer all the valid entailments. We shall not give this list here, the interested reader may consult [Belnap, 1977] or [Belnap, 1977a].

Instead we take into consideration another aspect which in a natural way is connected with the epistemic understanding of the truth degree set $\mathcal{W}_4^* = \{\mathbf{T}, \mathbf{F}, \mathbf{N}, \mathbf{B}\}$. The lattice ordering of $\mathcal{W}_4^*$, symbolized in Fig. 1, was based on the intuition that the "larger" elements are the better ones in the sense that the best possible case is definite truth, and the worst possible case is definite falsehood. Hence this lattice ordering $\leqq$ is a kind of *truth ordering.*

This truth ordering can be contrasted with a *knowledge ordering* which counts "complete" lack of knowledge of the truth value of some sentence φ as the worst case, and which prefers "more complete" knowledge. In some sense, this point of view transforms $\mathcal{W}_4^*$ into a set of knowledge degrees. From this point of view, another lattice ordering $\sqsubseteq$, and hence another lattice structure, becomes interesting which may be symbolized by the diagram of Fig. 3.

---

[12] This type of four-valued semantics may be used to give adequate semantical interpretations for different systems of relevance logic. We will not discuss details here. The interested reader may e.g. consult [Priest and Sylvan, 1992; Restall, 1995].

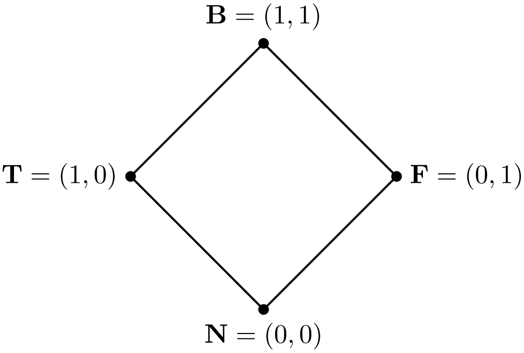


Figure 3. Four-valued knowledge degree lattice $\mathcal{W}_4^*$

Again here, the lattice ordering $\sqsubseteq$ goes "bottom-up". Of course, the diagram of Fig. 3 is just the diagram of Fig. 1 "turned around", which also means that the knowledge ordering $\sqsubseteq$ goes "left-to-right" in the diagram of Fig. 1.

From an algebraic point of view the truth degree set $\mathcal{W}_4^*$ with these two lattice structures becomes a particular case of a *bilattice*. This is a type of structure which more recently has been introduced in lattice theory, and which seems to be of particular interest for applications in logic and computer science, cf. e.g. [Baaz *et al.*, 1993; Fitting, 1988; Fitting, 1989; Ginsberg, 1987; Ginsberg, 1988].

6 LOGICS WITH T-NORM BASED CONNECTIVES

From now on we shall restrict our considerations to the case of infinite valued logics, even to logics with the real unit interval (3) as truth degree set. The main reason for this decision is that the recent development of many-valued logics has its focus on the development of such infinite valued systems. The core motivation for the emphasis on this type of systems is that they count as the background logics for the theory of fuzzy sets, as introduced by Zadeh [Zadeh, 1965]. And fuzzy sets have become an important tool in knowledge engineering as well as in artificial intelligence, hence are highly application relevant.

The fundamental infinite valued logics from §3 look quite different if one has in mind the form in which they first were presented.

Fortunately, however, there is a common generalization which allows to present all these three logics in a uniform way. In this uniform presentation one of the conjunction connectives becomes a core role: $\wedge$ in the system G_∞, & in the system L_∞, and $\odot$ in the system Π.

But this uniform generalization covers a much larger class of infinite valued logics over $[0,1]$: the core conjunction connective — which shall now in general be denoted & — has only to have a truth degree function $\otimes$ which, as a binary operation in the real unit interval, should be an associative, commutative, and isotonic operation which has 1 as a neutral element, i.e. should satisfy for arbitrary $x, y, z \in [0,1]$:

(T1) $x \otimes (y \otimes z) = (x \otimes y) \otimes z$,

(T2) $x \otimes y = y \otimes x$,

(T3) if $x \leq y$ then $x \otimes z \leq y \otimes z$,

(T4) $x \otimes 1 = x$.

Such binary operations are known as *t-norms* and have been used in the context of probabilistic metric spaces, cf. e.g. [Klement *et al.*, 2000]. At the same time they are considered as natural candidates for truth degree functions of conjunction connectives. And from such a t-norm one is able to derive (essentially) all the other truth degree functions for further connectives.

The minimum operation $u \wedge v$ from (8), the Łukasiewicz arithmetic conjunction $u \mathbin{\&} v$ from (12), and the ordinary product are the best known examples of t-norms.

In algebraic terms, such a t-norm $\otimes$ makes the real unit interval into an *ordered monoid*, i.e. into an abelian semigroup with unit element. And this ordered monoid is even *integral*, i.e. its unit element is at the same time the universal upper bound of the ordering. Additionally this monoid has because of

(27) $0 \otimes x \leq 0 \otimes 1 = 0$

the number 0 as an *annihilator*.

Starting from a t-norm $\otimes$ one finds a truth degree function $\rightarrowtail$ for an implication connective via the *adjointness condition*

(28) $x \otimes z \leq y \quad \Longleftrightarrow \quad z \leq (x \rightarrowtail y)\,.$

However, to guarantee that this adjointness condition (28) determines the operation $\rightarrowtail$ uniquely, one has to suppose that the t-norm $\otimes$ is a *left continuous* function in both arguments. Indeed, the adjointness condition (28) is equivalent to the condition that $\otimes$ is left continuous in both arguments, cf. [Gottwald, 2001].

Instead of this adjointness condition (28) one could equivalently either give the direct definition

(29) $x \rightarrowtail y = \sup\{z \mid x \otimes z \leq y\}$

of the *residuation operation* $\rightarrowtail$, or one could force the t-norm $\otimes$ to have the *sup-preservation property*

$$(30)\quad \sup_{i\to\infty}(x_i \otimes y) = (\sup_{i\to\infty} x_i) \otimes y$$

for each $y \in [0,1]$ and each non-decreasing sequence $(x_i)_{i\to\infty}$ from the real unit interval.

In this framework one additionally introduces a further unary operation $-$ by

$$(31)\quad -x =_{\text{df}} x \rightarrowtail 0\,,$$

and considers this as the truth degree function of a negation connective. That this works also in the formalized language of the corresponding system of logic forces to introduce into this language a truth degree constant $\overline{0}$ to denote the truth degree zero.

And finally one likes to have the weak conjunction and disjunction connectives $\wedge, \vee$ available. These connectives should also be added to the vocabulary. However, it suffices to add only the min-conjunction $\wedge$, because then for each left continuous t-norm $\otimes$ and its residuated implication $\rightarrowtail$ one has, completely similar to the situation (17) in the product logic,

$$(32)\quad u \vee v = ((u \rightarrowtail v) \rightarrowtail v) \wedge ((v \rightarrowtail u) \rightarrowtail u)\,.$$

All these considerations lead in a natural way to algebraic structures which, starting from the unit interval, consider a left continuous t-norm $\otimes$ together with its residuation operation $\rightarrowtail$, with the minimum-operation $\wedge$, and the maximum operation $\vee$ as basic operations of such an algebraic structure, and with the particular truth degrees $0, 1$ as fixed objects (i.e. as nullary operations) of the structure. Such an algebraic structure

$$(33)\quad \langle [0,1], \wedge, \vee, \otimes, \rightarrowtail, 0, 1\rangle$$

shall be coined to be a *t-norm algebra.*

7 RESIDUATED IMPLICATIONS VERSUS S-IMPLICATIONS

With the basic properties of classical logic in mind, particularly because of the logical equivalence of the formulas

$$(34)\quad \varphi \to \psi \qquad \text{and} \qquad \neg\varphi \vee \psi \qquad \text{and} \qquad \neg(\varphi \wedge \neg\psi)$$

in classical logic, the introduction of the implication connective directly as residuation via (29) or via the adjointness condition (28) seems to be quite sophisticated, and perhaps unnecessarily complicated.

If one has in mind an arrow-free approach similar to (34) for the present case, one would have to start for a definition of an implication connective either with a generalized disjunction or with a generalized conjunction, and had to add in any case a generalized negation.

In this case the implication, defined according to one of the equivalences in (34), is often coined "S-implication".

This has been done e.g. in [Butnariu *et al.*, 1995; Hekrdla *et al.*, 2003], but this approach does not really become simpler as the former one because one needs to fix either, besides the basic t-norm, an additional negation, or one has to fix a negation together with a disjunction.

However, the main disadvantage of such a modification is that one looses a quite natural strong soundness property one has for the rule of detachment in the former t-norm based approach. From the adjointness condition (28) one has always

(35) $u \otimes (u \rightarrowtail v) \leq v$

simply because of

$$u \otimes (u \rightarrowtail v) \leq v \quad \text{iff} \quad (u \rightarrowtail v) \otimes u \leq v \quad \text{iff} \quad u \rightarrowtail v \leq u \rightarrowtail v\,.$$

Therefore one has, in the t-norm based approach with implication defined as residuation, a natural lower bound for the truth degree of a formula ψ which has been derived from $\varphi \to \psi$ and φ via the rule of detachment: the truth degree of the formula $\varphi \,\&\, (\varphi \to \psi)$.

A similar property is lacking in general for the approach via S-implications. In this case, say starting from a t-norm $\otimes$ and a negation $\neg\!\!\!\rightharpoondown$, the corresponding property to (35) would be

(36) $u \otimes \neg\!\!\!\rightharpoondown (u \otimes \neg\!\!\!\rightharpoondown v) \leq v\,.$

But this fails already in the case that, independent of the choice of the t-norm $\otimes$, the negation $\neg\!\!\!\rightharpoondown$ is the Gödel negation $\sim$ of (9), i.e. the common negation of the systems G and Π. For this negation and any $v > 0$ one has $\sim v = 0$, hence $u \otimes \sim v = 0$, which means $\sim (u \otimes \sim v) = 1$ and $u \otimes \sim (u \otimes \sim v) = u$. Now one may choose $u > v$ to see that (36) fails.

8 CONTINUOUS T-NORMS

Among the large class of all t-norms the continuous ones are the best understood. A t-norm is continuous iff it is continuous as a real function of two variables, or equivalently, iff it is continuous in each argument (with the other one as a parameter), cf. [Gottwald, 2001; Klement *et al.*, 2000].

Furthermore, all continuous t-norms are ordinal sums of only three of them: the Łukasiewicz arithmetic t-norm $u \,\&\, v$ from (12), the ordinary product t-norm, and the minimum operation $u \wedge v$. The definition of an ordinal sum of t-norms is the following one.

DEFINITION 3. Suppose that $([a_i, b_i])_{i \in I}$ is a countable family of non-overlapping proper subintervals of the unit interval $[0, 1]$, let $(\boldsymbol{t}_i)_{i \in I}$ be a family of t-norms, and let $(\varphi_i)_{i\in I}$ be a family of mappings such that each φ_i is an order isomorphism from $[a_i, b_i]$ onto $[0,1]$. Then the (generalized) *ordinal sum* of the combined family $(([a_i, b_i], \boldsymbol{t}_i, \varphi_i))_{i\in I}$ is the binary function $T : [0,1]^2 \to [0,1]$ characterized by

$$(37)\quad T(u,v) = \begin{cases} {\varphi_k}^{-1}(\boldsymbol{t}_k(\varphi_k(u), \varphi_k(v)), & \text{if } u, v \in [a_k, b_k] \\ \min\{u, v\} & \text{otherwise.} \end{cases}$$

Often it is helpful to visualize the construction of an ordinal sum. For a simple case which shows some of the interval summands this is done in Fig. 4.

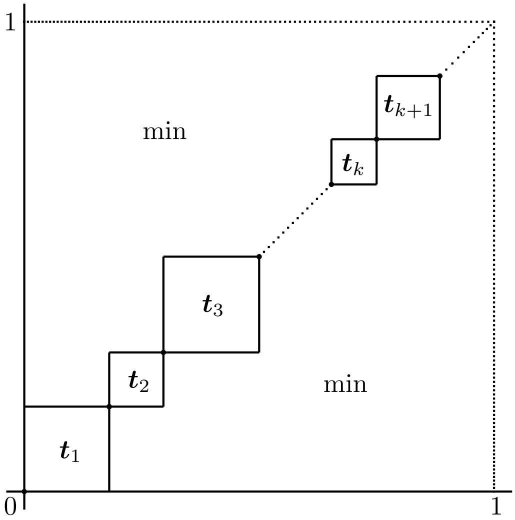


Figure 4. The basic construction of an ordinal sum

It is easy to see that an order isomorphic copy of the minimum t-norm is again the minimum operation. Thus the whole construction of ordinal sums of t-norms even allows to assume that the summands are formed from t-norms different from the minimum t-norm. This detail, however, shall be inessential for the present considerations.

But it should be mentioned that all the endpoints a_i, b_i of the interval family $([a_i, b_i])_{i\in I}$ give *idempotents* of the resulting ordinal sum t-norm T:

$$T(a_i, a_i) = a_i\,, \quad T(b_i, b_i) = b_i \quad \text{for all } i \in I.$$

Conversely, if one knows all the idempotents of a given continuous t-norm $\boldsymbol{t}$, i.e. all $u \in [0,1]$ with $\boldsymbol{t}(u,u) = u$, then one is able to give a representation of $\boldsymbol{t}$ as an ordinal sum, as explained again in [Klement *et al.*, 2000].

The general result, given e.g. in [Gottwald, 2001; Klement *et al.*, 2000], reads as follows.

THEOREM 4. *Each continuous t-norm* $\mathbf{t}$ *is the (generalized) ordinal sum of (isomorphic) copies of the Lukasiewicz t-norm, the product t-norm, and the minimum t-norm.*

As was mentioned in §3, the t-norm based logics which are determined by these three t-norms are well known and adequately axiomatized.

Therefore one is interested to find adequate axiomatizations also for further continuous t-norms. A global solution of this problem, i.e. a solution which did not only cover some few particular cases, appeared as quite difficult. Therefore, instead, one first has been interested to find all those formulas of the language of t-norm based systems which are logically valid in each one of these logics.

There seems to be a natural way to get an algebraic semantics for these considerations: the class of all t-norm algebras with a continuous t-norm should either form such an algebraic semantics, or should be a constitutive part — preferably a generating set — of a variety of algebraic structures which form such an algebraic semantics.

However, there seems to be an inadequacy in the description of this algebraic semantics: on the one hand the notion of t-norm algebra is a purely algebraic notion, the notion of continuity of a t-norm on the other hand is an analytical one. Fortunately, there is a possibility to give an algebraic characterization for the continuity of t-norms. It needs a further notion.

DEFINITION 5. A t-norm algebra $\langle [0,1], \wedge, \vee, \otimes, \rightarrowtail, 0, 1\rangle$ is *divisible* iff one has for all $a, b \in L$:

(38) $a \wedge b = a \otimes (a \rightarrowtail b)\,.$

And this notion gives the algebraic counterpart for the continuity, as shown e.g. in [Gottwald, 2001; Klement *et al.*, 2000].

PROPOSITION 6. *A t-norm algebra* $\langle [0,1], \wedge, \vee, \otimes, \rightarrowtail, 0, 1\rangle$ *is divisible iff the t-norm* $\otimes$ *is continuous.*

9 THE LOGIC OF CONTINUOUS T-NORMS

The class of t-norm algebras (with a continuous t-norm or not) is not a variety: it is not closed under direct products because each t-norm algebra is linearly ordered, but the direct products of linearly ordered structures are not linearly ordered, in general. Hence one may expect that it would be helpful for the development of a logic of continuous t-norms to extend the class of all divisible t-norm algebras in a moderate way to get a variety. And indeed this idea works: it was developed by P. Hájek and in detail explained in [Hájek, 1998].

The core points are that one considers instead of the divisible t-norm algebras, which are linearly ordered integral monoids as mentioned previously, now lattice ordered integral monoids which are divisible, which have an additional residuation operation connected with the semigroup operation via an adjointness condition like (28), and which satisfy a pre-linearity condition like (18). These structures have been called BL-algebras; they are completely defined in the following way.

DEFINITION 7. A *BL-algebra* is an algebraic structure

(39) $\mathbf{L} = \langle L, \vee, \wedge, *, \rightarrow, \mathbf{0}, \mathbf{1} \rangle$

with four binary operations and two constants such that

(*i*) $(L, \vee, \wedge, \mathbf{0}, \mathbf{1})$ is a bounded lattice, i.e. has $\mathbf{0}$ and $\mathbf{1}$ as the universal lower and upper bounds w.r.t. the lattice ordering $\leq$,

(*ii*) $(L, *, \mathbf{1})$ is an abelian monoid, i.e. a commutative semigroup with unit $\mathbf{1}$ such that the multiplication $*$ is associative, commutative and satisfies $\mathbf{1} * x = x$ for all $x \in L$,

(*iii*) the binary operations $*$ and $\rightarrow$ form an adjoint pair, i.e. satisfy for all $x, y, z \in L$ the *adjointness* condition

(40) $z \leq (x \rightarrow y) \iff x * z \leq y$

(*iv*) and moreover, for all $x, y \in L$ one has satisfied the *pre-linearity* condition

(41) $(x \rightarrow y) \vee (y \rightarrow x) = \mathbf{1}$

as well as the *divisibility* condition

(42) $x * (x \rightarrow y) = x \wedge y\,.$

The axiomatization of Hájek [1998] for the *basic t-norm logic* BL (in [Gottwald, 2001] denoted BTL), i.e. for the class of all well-formed formulas which are valid in all BL-algebras, is given in a language $\mathcal{L}_T$ which has as basic vocabulary the connectives $\rightarrow, \&$ and the truth degree constant $\overline{0}$, taken in each BL-algebra $\langle L, \cap, \cup, *, \rightarrowtail, 0, 1 \rangle$ as the operations $\rightarrowtail, *$ and the element 0. Then this t-norm based logic has as axiom system $\mathsf{Ax}_{\mathsf{BL}}$ the following schemata:

($\mathsf{Ax}_{\mathsf{BL}}1$) $(\varphi \rightarrow \psi) \rightarrow ((\psi \rightarrow \chi) \rightarrow (\varphi \rightarrow \chi))\,,$

($\mathsf{Ax}_{\mathsf{BL}}2$) $\varphi \,\&\, \psi \rightarrow \varphi\,,$

($\mathsf{Ax}_{\mathsf{BL}}3$) $\varphi \,\&\, \psi \rightarrow \psi \,\&\, \varphi\,,$

($\mathsf{Ax}_{\mathsf{BL}}4$) $(\varphi \rightarrow (\psi \rightarrow \chi)) \rightarrow (\varphi \,\&\, \psi \rightarrow \chi)\,,$

($\mathsf{Ax}_{\mathsf{BL}}5$) $(\varphi \,\&\, \psi \rightarrow \chi) \rightarrow (\varphi \rightarrow (\psi \rightarrow \chi))\,,$

($\mathsf{Ax}_{\mathsf{BL}}6$) $\varphi \,\&\, (\varphi \rightarrow \psi) \rightarrow \psi \,\&\, (\psi \rightarrow \varphi)\,,$

($\mathsf{Ax}_{\mathsf{BL}}7$) $((\varphi \rightarrow \psi) \rightarrow \chi) \rightarrow (((\psi \rightarrow \varphi) \rightarrow \chi) \rightarrow \chi)\,,$

($\mathsf{Ax}_{\mathsf{BL}}8$) $\overline{0} \rightarrow \varphi\,,$

and has as its (only) inference rule the rule of detachment, or: modus ponens (w.r.t. the implication connective →).

The logical calculus which is constituted by this axiom system and its inference rule, and which has the standard notion of derivation, shall be denoted by $\mathbb{K}_{\mathsf{BL}}$ or just by BL. (Similarly in other cases.)

Starting from the primitive connectives $\rightarrow, \&$ and the truth degree constant $\overline{0}$, the language $\mathcal{L}_T$ of BL is extended by definitions of additional connectives $\wedge, \vee, \neg$:

$$\varphi \wedge \psi \quad =_{\mathrm{df}} \quad \varphi \,\&\, (\varphi \rightarrow \psi)\,, \tag{43}$$

$$\varphi \vee \psi \quad =_{\mathrm{df}} \quad ((\varphi \rightarrow \psi) \rightarrow \psi) \wedge ((\psi \rightarrow \varphi) \rightarrow \varphi)\,, \tag{44}$$

$$\neg\varphi \quad =_{\mathrm{df}} \quad \varphi \rightarrow \overline{0}\,, \tag{45}$$

where φ, ψ are formulas of the language of that system.

Calculations (in BL-algebras) show that the additional connectives $\wedge, \vee$ just have the lattice operations $\cap, \cup$ as their truth degree functions.

It is a routine matter, but a bit tedious, to check that this logical calculus $\mathbb{K}_{\mathsf{BL}}$, usually called the axiomatic system BL, is sound, i.e. derives only such formulas which are valid in all BL-algebras. A proof is given in [Hájek, 1998], together with a proof of a corresponding completeness theorem.

COROLLARY 8. *The Lindenbaum algebra of the axiomatic system* BL *is a BL-algebra.*

THEOREM 9 (General Completeness). *A formula* φ *of the language* $\mathcal{L}_T$ *is derivable within the axiomatic system* BL *iff* φ *is valid in all BL-algebras.*

The proof method yields that each BL-algebra is (isomorphic to) a subdirect product of linearly ordered BL-algebras, i.e. of BL-chains. Thus it allows a nice modification of the previous result.

COROLLARY 10 (General Completeness; Version 2). *A formula* φ *of* $\mathcal{L}_T$ *is derivable within the axiomatic system* BL *iff* φ *is valid in all BL-chains.*

But even more is provable and leads back to the starting point of the whole approach: the logical calculus $\mathbb{K}_{\mathsf{BL}}$ characterizes just those formulas which hold true w.r.t. all divisible t-norm algebras. This was proved in [Cignoli *et al.*, 2000].

THEOREM 11 (Standard Completeness). *The class of all formula which are provable in the system* BL *coincides with the class of all formulas which are logically valid in all t-norm algebras with a continuous t-norm.*

The main steps in the proof are to show (i) that each BL-algebra is a subdirect product of subdirectly irreducible BL-chains, i.e. of linearly ordered BL-algebras which are not subdirect products of other BL-chains, and (ii) that each subdirectly irreducible BL-chain can be embedded into the ordinal sum of some BL-chains which are either trivial one-element BL-chains, or linearly ordered MV-algebras, or linearly ordered product algebras, such that (iii) each such ordinal summand is locally embedable into a t-norm based residuated lattice with a continuous t-norm, cf. [Cignoli *et al.*, 2000; Hájek, 1998] and again [Gottwald, 2001].

This is a lot more of algebraic machinery as necessary for the proof of the General Completeness Theorem 9 and thus offers a further indication that the extension of the class of divisible t-norm algebras to the class of BL-algebras made the development of the intended logical system easier. But even more can be seen from this proof: *the class of BL-algebras is the smallest variety which contains all the divisible t-norm algebras.* And the algebraic reason for this is that each variety may be generated from its subdirectly irreducible elements, cf. again [Burris and Sankappanavar, 1981; Denecke and Wismath, 2002; Grätzer, 1979].

And another generalization of Theorem 9 deserves to be mentioned. To state it, let us call *schematic extension* of BL every extension which consists in an addition of finitely many axiom schemata to the axiom schemata of BL. And let us denote such an extension by $\mathsf{BL}(\mathcal{C})$. And call BL($\mathcal{C}$)-algebra each BL-algebra $\mathbf{A}$ which makes $\mathbf{A}$-valid all formulas of $\mathcal{C}$.

Then one can prove, as done in [Hájek, 1998], an even more general completeness result.

THEOREM 12 (Extended General Completeness). *For each finite set $\mathcal{C}$ of axiom schemata and any formula φ of $\mathcal{L}_T$ there are equivalent:*

(i) φ is derivable within $\mathsf{BL}(\mathcal{C})$;

(ii) φ is valid in all $\mathsf{BL}(\mathcal{C})$-algebras;

(iii) φ is valid in all $\mathsf{BL}(\mathcal{C})$-chains.

The extension of these considerations to the first-order case is also given in [Hájek, 1998], but shall not be discussed here.

But the algebraic machinery allows even deeper insights. After some particular results e.g. in [Haniková, 2001; Haniková, 2003], the study of such subvarieties of the variety of all BL-algebras which are generated by single t-norm algebras $\langle [0,1], \wedge, \vee, \otimes, \rightarrowtail, 0, 1\rangle$ with a continuous t-norm $\otimes$ led to (finite) axiomatizations of those t-norm based logics which have a standard semantics determined just by this continuous t-norm algebra. These results have recently been presented in [Esteva *et al.*, 2004].

10 THE LOGIC OF LEFT CONTINUOUS T-NORMS

The guess of Esteva/Godo [Esteva and Godo, 2001] has been that one should arrive at the logic of left continuous t-norms if one starts from the logic of continuous t-norms and deletes the continuity condition, i.e. the divisibility condition (38).

The algebraic approach needs only a small modification: in Definition 7 of BL-algebras one has simply to delete the divisibility condition (42). The resulting algebraic structures have been called *MTL-algebras*. They again form a variety.

Following this idea, one has to modify the previous axiom system in a suitable way. And one has to delete the definition (43) of the connective $\wedge$, because this definition (together with suitable axioms) essentially codes the divisibility condition. The definition (44) of the connective $\vee$ remains unchanged.

As a result one now considers a new system MTL of mathematical fuzzy logic, characterized semantically by the class of all MTL-algebras. It is connected with the axiom system

$(\mathsf{Ax_{MTL}}1)$ $\quad (\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))\,,$

$(\mathsf{Ax_{MTL}}2)$ $\quad \varphi \,\&\, \psi \to \varphi\,,$

$(\mathsf{Ax_{MTL}}3)$ $\quad \varphi \,\&\, \psi \to \psi \,\&\, \varphi\,,$

$(\mathsf{Ax_{MTL}}4)$ $\quad (\varphi \to (\psi \to \chi)) \to (\varphi \,\&\, \psi \to \chi)\,,$

$(\mathsf{Ax_{MTL}}5)$ $\quad (\varphi \,\&\, \psi \to \chi) \to (\varphi \to (\psi \to \chi))\,,$

$(\mathsf{Ax_{MTL}}6)$ $\quad \varphi \wedge \psi \to \varphi\,,$

$(\mathsf{Ax_{MTL}}7)$ $\quad \varphi \wedge \psi \to \psi \wedge \varphi\,,$

$(\mathsf{Ax_{MTL}}8)$ $\quad \varphi \,\&\, (\varphi \to \psi) \to \varphi \wedge \psi\,,$

$(\mathsf{Ax_{MTL}}9)$ $\quad \overline{0} \to \varphi\,,$

$(\mathsf{Ax_{MTL}}10)$ $\quad ((\varphi \to \psi) \to \chi) \to (((\psi \to \varphi) \to \chi) \to \chi)\,,$

together with the rule of detachment (w.r.t. the implication connective →) as (the only) inference rule.

It is a routine matter, but again tedious, to check that this logical calculus $\mathbb{K}_{\mathsf{MTL}}$ is sound, i.e. derives only such formulas which are valid in all MTL-algebras.

COROLLARY 13. *The Lindenbaum algebra of the logical calculus* $\mathbb{K}_{\mathsf{MTL}}$ *is an MTL-algebra.*

Proofs of this result and also of the following completeness theorem are given in [Esteva and Godo, 2001].

THEOREM 14 (General Completeness). *A formula* φ *of the language* $\mathcal{L}_T$ *is derivable within the logical calculus* $\mathbb{K}_{\mathsf{MTL}}$ *iff* φ *is valid in all MTL-algebras.*

Again the proof method yields that each MTL-algebra is (isomorphic to) a subdirect product of linearly ordered MTL-algebras, i.e. of MTL-chains.

COROLLARY 15 (General Completeness; Version 2). *A formula* φ *of* $\mathcal{L}_T$ *is derivable within the axiomatic system* MTL *iff* φ *is valid in all MTL-chains.*

And again, similar as for the BL-case, even more is provable: the logical calculus $\mathbb{K}_{\mathsf{MTL}}$ characterizes just these formulas which hold true w.r.t. all those t-norm based logics which are determined by a left continuous t-norm. A proof is given in [Jenei and Montagna, 2002].

THEOREM 16 (Standard Completeness). *The class of all formulas which are provable in the logical calculus* $\mathbb{K}_{\mathsf{MTL}}$ *coincides with the class of all formulas which are logically valid in all t-norm algebras with a left continuous t-norm.*

This result again means, as the similar one for the logic of continuous t-norms, that the variety of all MTL-algebras is the smallest variety which contains all t-norm algebras with a left continuous t-norm.

Because of the fact that the BL-algebras are the divisible MTL-algebras, one gets another adequate axiomatization of the basic t-norm logic BL if one extends the axiom system $\mathbb{K}_{\mathsf{MTL}}$ with the additional axiom schema

(46) $\varphi \wedge \psi \to \varphi \mathbin{\&} (\varphi \to \psi)\,.$

The simplest way to prove that this implication is sufficient is to show that the inequality $x * (x \rightarrowtail y) \leq x \cap y$, which corresponds to the converse implication, holds true in each MTL-algebra. Similar remarks apply to further extensions of MTL we are going to mention.

Also for MTL an extended completeness theorem similar to Theorem 12 remains true.

THEOREM 17 (Extended General Completeness). *For each finite set* $\mathcal{C}$ *of axiom schemata and any formula* φ *of* $\mathcal{L}_T$ *the following are equivalent:*

(*i*) φ *is derivable within the logical calculus* $\mathbb{K}_{\mathsf{MTL}} + \mathcal{C}$*;*

(*ii*) φ *is valid in all MTL*($\mathcal{C}$)*-algebras;*

(*iii*) φ *is valid in all MTL*($\mathcal{C}$)*-chains.*

Again the extension to the first-order case is similar to the treatment in [Hájek, 1998] for BL and shall not be discussed here.

11 SOME GENERALIZATIONS

The standard approach toward t-norm based logics, as explained in §§9 and 10, has been modified in various ways. The main background ideas are the extension or the modification of the expressive power of these logical systems.

A first, quite fundamental addition to the standard vocabulary of the languages of t-norm based systems was proposed in [Baaz, 1996]: a unary propositional operator $\triangle$ which has for t-norm algebras the semantics

(47) $\triangle(x) = 1 \quad \text{for } x = 1\,, \qquad \triangle(x) = 0 \quad \text{for } x \neq 1\,.$

This unary connective can be added to the systems BL and MTL via the additional axioms

(Δ1) $\triangle\varphi \vee \neg\triangle\varphi$

(Δ2) $\triangle(\varphi \vee \varphi) \to (\triangle\varphi \vee \triangle\psi)$

(Δ3) $\triangle\varphi \to \varphi$

(Δ4) $\triangle\varphi \to \triangle\triangle\varphi$

(Δ5) $\triangle(\varphi \to \psi) \to (\triangle\varphi \to \triangle\psi)$

This addition leaves all the essential theoretical results, like correctness and completeness theorems, valid: of course w.r.t. suitably expanded algebraic structures.

A second stream of papers discusses the addition of an idempotent negation, i.e. a negation which satisfies the double negation law, for those cases where the standard negation of the t-norm based system is not idempotent. This is e.g. the case for the product logic which, as explained at the end of Subsection 3.3, has the Gödel negation (9) as its standard negation. By the way, it should be noticed that (routine calculations show that) this non-idempotent Gödel negation is the standard negation of all those t-norm algebras with a t-norm $\otimes$ which does not have zero-divisors.[13] A very general approach is given in [Esteva *et al.*, 2000], and a more particular axiomatization problem discussed in [Gottwald and Jenei, 2001].

Another stream of papers, partly related to the previously mentioned one, is devoted to the problem of a unified treatment of different, usually two, t-norms and their related connectives within one logical system. Here the focus is on the join of the systems based upon the Łukasiewicz t-norm and upon the product t-norm. The great advantage of this unification is that the Łukasiewicz t-norm essentially allows to treat the addition, as may be seen from the truth degree function (12) of the Łukasiewicz (arithmetical) disjunction, and that the product t-norm adds the treatment of the usual product: and this means that the elementary arithmetic (in the unit interval) can be discussed in this combined system. This combined system has been considered in two strongly related forms, denoted ŁΠ and ŁΠ$\frac{1}{2}$. The distinction between both systems is that ŁΠ has both t-norms & and $\odot$ and their related (residual) implications and negations among their basic connectives, and that ŁΠ$\frac{1}{2}$ adds a truth degree constant for the truth degree $\frac{1}{2}$. These two systems are discussed in detail in [Cintual, 2001; Cintula, 2001a; Cintula, 2001b; Esteva and Godo, 1999; Esteva *et al.*, 2001].

A fourth stream of papers intends to weaken the systems BL and MTL in such a way that one deletes the explicit reference to the truth degree constant $\overline{0}$ and considers the *falsity free* fragments of the previous systems. From the algebraic point of view their characteristic structures become the *hoops* which in general are defined as algebraic structures $H = \langle H, *, \Rightarrow, \mathbf{1}\rangle$ such that $\langle H, *, \mathbf{1}\rangle$ is an abelian monoid and that the further binary operation $\Rightarrow$ satisfies the equations

$$\begin{aligned} x \Rightarrow x &= \mathbf{1}\,, \\ x*(x \Rightarrow y) &= y*(y \Rightarrow x)\,, \\ (x*y) \Rightarrow z &= x \Rightarrow (y \Rightarrow z)\,. \end{aligned}$$

The definition

$$x \sqsubseteq y \;=_{\text{def}}\; x \Rightarrow x = \mathbf{1}$$

---

[13] Zero-divisors of a t-norm $\boldsymbol{t}$ are such reals $0 < u, v < 1$ for which $\boldsymbol{t}(u, v) = 0$ holds.

provides an ordering $\sqsubseteq$ with universal upper bound $\mathbf{1}$ which makes $\langle H, *, \mathbf{1}\rangle$ an ordered monoid, and which has the additional property that the operations $*, \Rightarrow$ become an adjoint pair w.r.t. this ordering.

In particular, hoops with the additional property

$$x \Rightarrow (y \Rightarrow z) \ \sqsubseteq\ (y \Rightarrow (x \Rightarrow z)) \Rightarrow z$$

can in a natural way be generated from t-norm algebras with continuous t-norms, as has been shown in [Agliano *et al.*, 200X]. So one has a kind of competing generalization of t-norm algebras. And for this kind of algebraic semantics one can find adequate axiomatizations for corresponding hoop logics quite similar to the approaches of §§9 and 10. The details have been developed in [Esteva *et al.*, 2003].

And a fifth stream discusses the generalization of the algebraic semantics from the case of abelian lattice ordered monoids with residuation to the case of non-commutative lattice ordered semigroups. In this context one tries to define non-commutative BL-algebras or non-commutative MTL-algebras, and similarly defines non-commutative t-norms, also called pseudo-t-norms. And these considerations become combined with the design of an adequate axiomatization, with similar results as in §§9 and 10. The most important ones of these papers are [di Nola *et al.*, 2002; Flondor *et al.*, 2001; Hájek, 2003; Hájek, 2003a; Hájek, 2003b; Jenei and Montagna, 2003; Kühr, 2003].

And finally it should be mentioned that Hájek

[2005] even gives a common generalization of all of these generalized fuzzy logics, thus giving up divisibility, the falsity constant, and commutativity. The corresponding algebras are called *fleas* (or *flea algebras*), and the logic is the *flea logic* FlL. There are examples of fleas on $(0, 1]$ not satisfying divisibility, nor commutativity, and having no least element.

It shall be sufficient to mention these generalizations here. The interested reader can find more details in the survey paper [Gottwald and Hájek, 2005] and in the original publications.

12 PAVELKA STYLE EXTENSIONS

Having in mind that fuzzy logics, also in their form as formalized logical systems, should be a (mathematical) tool for approximative reasoning makes it desirable that they should be able to deal with graded inferences.

The systems of t-norm based logics discussed up to now have been designed to formalize the logical background for fuzzy sets, and they allow themselves for degrees of truth of their formulas. But they all have crisp notions of consequence, i.e. of entailment and of provability.

It is natural to ask whether it is possible to generalize these considerations to the case that one starts from *fuzzy sets of formulas*, and that one gets from them as consequence hulls again fuzzy sets of formulas. This problem was first treated by Pavelka [1979]. The basic monograph elaborating this approach is [Novák *et al.*, 1999]. We discuss in the present section this kind of approach, because it uses graded relations of entailment and of provability.

However, it should be mentioned that there is also another, more algebraically oriented approach toward consequence operations for the classical case, originating from Tarski [1930] and presented e.g. in [Wójcicki, 1988]. This approach treats consequence operations as closure operations. And this type of approach has been generalized to closure operations in classes of fuzzy sets of formulas by Gerla [2001]. It shall be discussed in §13.

The Pavelka-style approach has to deal with fuzzy sets $\Sigma^{\sim}$ of formulas, i.e. besides formulas φ also their membership degrees $\Sigma^{\sim}(\varphi)$ in $\Sigma^{\sim}$. And these membership degrees are just the truth degrees. We may assume that these degrees again form a residuated lattice $\mathbf{L} = \langle L, \cap, \cup, *, \rightarrowtail, 0, 1\rangle$. Thus we (slightly) generalize the standard notion of fuzzy set (with membership degrees from the real unit interval). Therefore the appropriate language has the same logical connectives as in the previous considerations.

The Pavelka-style approach is an easy matter as long as the entailment relationship is considered. An evaluation e is a *model* of a fuzzy set $\Sigma^{\sim}$ of formulas iff

$$(48)\quad \Sigma^{\sim}(\varphi) \leqslant \mathsf{Val}(\varphi, e)$$

holds for each formula φ. This immediately yields as definition of the entailment relation that the semantic consequence hull of $\Sigma^{\sim}$ should be characterized by the membership degrees

$$(49)\quad \mathcal{C}^{\mathsf{sem}}(\Sigma^{\sim})(\psi) = \bigwedge\{\mathsf{Val}(\psi, e) \mid e \text{ model of } \Sigma^{\sim}\}$$

for each formula ψ.

For a syntactic characterization of this entailment relation it is necessary to have some calculus $\mathbb{K}$ which treats formulas of the language together with truth degrees. So the language of this calculus has to extend the language of the basic logical system by having also symbols for the truth degrees. Depending upon the truth degree structure, this may mean that the language of this calculus becomes an uncountable one.

Further on we indicate these symbols by overlined letters like $\overline{a}, \overline{c}$. And we realize the common treatment of formulas and truth degrees by considering *evaluated formulas*, i.e. ordered pairs $(\overline{a}, \varphi)$ consisting of a truth degree symbol and a formula. This trick transforms in a natural way each fuzzy set $\Sigma^{\sim}$ of formulas into a (crisp) set of evaluated formulas, again denoted by $\Sigma^{\sim}$.

So $\mathbb{K}$ has to allow to derive evaluated formulas out of sets of evaluated formulas, of course using suitable axioms and rules of inference. These axioms are usually only formulas φ which, however, are used in the derivations as the corresponding evaluated formulas $(\overline{1}, \varphi)$. Derivations in $\mathbb{K}$ out of some set $\Sigma^{\sim}$ of evaluated formulas are finite sequences of evaluated formulas which either are axioms, or elements of (the support of) $\Sigma^{\sim}$, or result from former evaluated formulas by application of one of the inference rules.

Each $\mathbb{K}$-derivation of an evaluated formula $(\overline{a}, \varphi)$ counts as a derivation of φ *to the degree* $a \in L$. The *provability degree* of φ from $\Sigma^{\sim}$ in $\mathbb{K}$ is the supremum over all these degrees. This now yields that the syntactic consequence hull of $\Sigma^{\sim}$ should be the fuzzy set $\mathcal{C}^{\mathsf{syn}}_{\mathbb{K}}$ of formulas characterized by the membership function

$$(50)\quad \mathcal{C}^{\mathsf{syn}}_{\mathbb{K}}(\Sigma^{\sim})(\psi) = \bigvee\{a \in L \mid \mathbb{K} \text{ derives } (\overline{a}, \psi) \text{ out of } \Sigma^{\sim}\}$$

for each formula ψ.

Despite the fact that $\mathbb{K}$ is a standard calculus, this is an *infinitary* notion of provability.

For the infinite-valued Łukasiewicz logic L this machinery works particularly well because it needs in an essential way the continuity of the residuation operation. In this case we can form a calculus $\mathbb{K}_{\mathsf{L}}$ which gives an adequate axiomatization for the graded notion of entailment in the sense that one has suitable soundness and completeness results.

This calculus $\mathbb{K}_{\mathsf{L}}$ has as axioms any axiom system of the infinite-valued Łukasiewicz logic L which provides together with the rule of detachment an adequate axiomatization of L, but $\mathbb{K}_{\mathsf{L}}$ replaces this standard rule of detachment by the generalized form

$$(51)\quad \frac{(\overline{a}, \varphi) \qquad (\overline{c}, \varphi \to \psi)}{(\overline{a * c}, \psi)}$$

for evaluated formulas.

The soundness result for this calculus $\mathbb{K}_{\mathsf{L}}$ yields the fact that the $\mathbb{K}_{\mathsf{L}}$-provability of an evaluated formula $(\overline{a}, \varphi)$ says that $a \leq \mathsf{Val}(\varphi, e)$ holds for every valuation e, i.e. that the formula $\overline{a} \to \varphi$ is valid — however as a formula of an *extended* propositional language which has all the truth degree constants among its vocabulary. Of course, now the evaluations e have also to satisfy $e(\overline{a}) = a$ for each $a \in [0, 1]$.

And the soundness and completeness results for $\mathbb{K}_{\mathsf{L}}$ say that a *strong completeness theorem* holds true giving

$$(52)\quad \mathcal{C}^{\mathsf{sem}}(\Sigma^{\sim})(\psi) \;=\; \mathcal{C}^{\mathsf{syn}}_{\mathbb{K}_{\mathsf{L}}}(\Sigma^{\sim})(\psi)$$

for each formula ψ and each fuzzy set $\Sigma^{\sim}$ of formulas.

If one takes the previously mentioned turn and extends the standard language of propositional L by truth degree constants for all degrees $a \in [0, 1]$, and if one reads each evaluated formula $(\overline{a}, \varphi)$ as the formula $\overline{a} \to \varphi$, then a slight modification $\mathbb{K}^{+}_{\mathsf{L}}$ of the former calculus $\mathbb{K}_{\mathsf{L}}$ again provides an adequate axiomatization: one has to add the *bookkeeping axioms*

$$\begin{aligned}(\overline{a} \mathbin{\&} \overline{c}) &\equiv \overline{a * c}\,,\\ (\overline{a} \to \overline{c}) &\equiv \overline{a \rightarrowtail_{\mathsf{L}} c}\,,\end{aligned}$$

as explained e.g. in [Novák *et al.*, 1999]. And if one is interested to have evaluated formulas together with the extension of the language by truth degree constants, one has also to add the *degree introduction rule*

$$\frac{(\overline{a}, \varphi)}{\overline{a} \to \varphi} .$$

However, even a stronger result is available which refers only to a notion of derivability over a countable language. The completeness result (52), for $\mathbb{K}_{\mathsf{L}}^{+}$ instead of $\mathbb{K}_{\mathsf{L}}$, becomes already provable if one adds truth degree constants only for all the *rationals* in $[0,1]$, as was shown in [Hájek, 1998]. And this extension of L is even only a conservative one, cf. [Hájek *et al.*, 2000], i.e. $\mathbb{K}_{\mathsf{L}}^{+}$ proves only such *constant-free* formulas of the language with rational constants which are already provable in the standard infinite-valued Łukasiewicz logic L.

For more details the reader may also consult e.g. [Hájek, 1998; Novák *et al.*, 1999; Turunen, 1995].

13 GERLA'S GENERAL APPROACH

For completeness we mention also a much more abstract approach toward fuzzy logics with graded notions of entailment as the previously explained one for the t-norm based fuzzy logics is.

The background for this generalization by G. Gerla, in detail explained in [Gerla, 2001], is that (already) in systems of classical logic the syntactic as well as the semantic consequence relations, i.e. the provability as well as the entailment relations, are closure operators within the set of formulas. This is a fundamental observation made by Tarski [1930] already in 1930. And the same holds true for the Pavelka style extensions of §12 and the operators $\mathcal{C}^{\mathsf{sem}}$ and $\mathcal{C}^{\mathsf{syn}}$ introduced in (49) and (50), respectively: they are generalized closure operators.

The context, chosen in [Gerla, 2001], is that of L-fuzzy sets, with $\mathbf{L} = \langle L, \leqslant \rangle$ an arbitrary complete lattice. A *closure operator in* $\mathbf{L}$ is a mapping $J : L \to L$ satisfying for arbitrary $x, y \in L$ the well known conditions

$$\begin{aligned} & x \leqslant J(x)\,, && \text{(increasingness)} \\ & x \leqslant y \Rightarrow J(x) \leqslant J(y)\,, && \text{(isotonicity)} \\ & J(J(x) = J(x)\,. && \text{(idempotency)} \end{aligned}$$

And a *closure system* in $\mathbf{L}$ is a subclass $C \subseteq L$ which is closed under arbitrary lattice meets.

For fuzzy logic such closure operators and closure systems are considered in the lattice $\mathcal{F}_L(\mathbb{F})$ of all fuzzy subsets of the set $\mathbb{F}$ of formulas of some suitable formalized language.

An *abstract fuzzy deduction system* now is an ordered pair $\mathcal{D} = (\mathcal{F}_L(\mathbb{F}), D)$ determined by a closure operator D in the lattice $\mathcal{F}_L(\mathbb{F})$. And the *fuzzy theories*

T of such an abstract fuzzy deduction system, also called $\mathcal{D}$-theories, are the fixed points of D: $T = D(T)$, i.e. the deductively closed fuzzy sets of formulas.

A rather abstract setting is also chosen for the semantics of such an abstract fuzzy deduction system: an *abstract fuzzy semantics* $\mathcal{M}$ is nothing but a class of elements of the lattice $\mathcal{F}_L(\mathbb{F})$, i.e. a class of fuzzy sets of formulas. These fuzzy sets of formulas are called *models*. The only restriction is that the universal set over $\mathbb{F}$, i.e. the fuzzy subset of $\mathbb{F}$ which has always membership degree one, is not allowed as a model. The background idea here is that, for each standard interpretation $\mathfrak{A}$ (in the sense of many-valued logic — including an evaluation of the individual variables) for the formulas of $\mathbb{F}$, a model M is determined as the fuzzy set which has for each formula $\varphi \in \mathbb{F}$ the truth degree of φ in $\mathfrak{A}$ as membership degree. Accordingly the satisfaction relation $\models_{\mathcal{M}}$ coincides with inclusion: for models $M \in \mathcal{M}$ and fuzzy sets Σ of formulas one has:

(53) $M \models_{\mathcal{M}} \Sigma \;\Leftrightarrow\; \Sigma \subseteq M\,.$

In this setting, one has a semantic and a syntactic consequence operator, both being closure operators, i.e. one has for each fuzzy set Σ of formulas from $\mathbb{F}$ a semantic as well as a syntactic consequence hull, given by

(54) $\mathcal{C}^{\mathsf{sem}}(\Sigma) = \bigcap\{M \in \mathcal{M} \mid M \models_{\mathcal{M}} \Sigma\}\,, \qquad \mathcal{C}^{\mathsf{syn}}(\Sigma) = D(\Sigma)\,.$

Similar to the classical case one has $\mathcal{C}^{\mathsf{sem}}(M) = M$ for each model $M \in \mathcal{M}$, i.e. each such model provides a $\mathcal{C}^{\mathsf{sem}}$-theory.

However, a general completeness theorem is not available. What one needs instead, in search for a completeness result, that are specifications which restrict the full generality of this approach, and lead mainly back to situations which have been discussed in the previous sections.

14 SOME RECENT APPLICATIONS

14.1 Fuzzy sets theory

It is an old approach, dating back to the early days of fuzzy set theory, to identify the membership degrees of fuzzy sets with truth degrees of a suitable many-valued logic. In different forms, this idea has been offered and explained e.g. in [Giles, 1976; Goguen, 1968; Gottwald, 1984; Gottwald, 1993; Klaua, 1966]. And it has since been the topic of occasional investigations like in [Takeuti and Titani, 1984; Takeuti and Titani, 1992].

This point of view toward fuzzy set theory has been one of the motivations behind the development of mathematical fuzzy logics. Therefore one may expect that the recent results in this field of mathematical fuzzy logics give rise to a return to this starting point to use the new insights e.g. for a coherent development of a (formalized) fuzzy set theory.

Indeed, the paper [Hájek and Haniková, 2003] and the subsequent Ph.D. Thesis [Haniková, 2003] use the (first-order) logic BL of continuous t-norms, extended with the $\triangle$-operator mentioned in (47), to develop a ZF-like axiomatization for a formalized fuzzy set theory together with a kind of standard model constructed in the style of Boolean valued models for (standard) set theory, as explained e.g. in [Bell, 1985].

The axioms are suitable versions of the axioms of extensionality, pairing, union, powerset, $\in$-induction (i.e. foundation), separation, collection (i.e. comprehension), and infinity, together with an axiom stating the existence of the *support* of each fuzzy set.

The standard model for this theory is formed w.r.t. some complete BL-chain $\mathbf{L} = \langle L, \wedge, \vee, *, \rightarrow, 0, 1\rangle$ and given by the transfinite hierarchy

$$(55)\quad V_0^{\mathbf{L}} = \emptyset\,, \qquad V_{\alpha+1}^{\mathbf{L}} = \left\{ f \in {}^{\mathsf{dom}(u)}L \mid \mathsf{dom}(u) \subseteq V_\alpha^{\mathbf{L}} \right\},$$

with unions at limit stages.[14]

The primitive predicates $\in, \subseteqq, =$ are interpreted using the following definitions for their truth degrees $[\![\ldots]\!]$:

$$\begin{aligned} [\![x \in y]\!] &= \bigcup_{u \in \mathsf{dom}(y)} ([\![u = x]\!] * y(u))\,, \\ [\![x \subseteqq y]\!] &= \bigcap_{u \in \mathsf{dom}(x)} (x(u) \Rightarrow [\![u \in y]\!])\,, \\ [\![x = y]\!] &= \triangle[\![x \subseteqq y]\!] * \triangle[\![y \subseteqq x]\!]\,. \end{aligned}$$

The last condition forces the equality to be crisp.

Besides this “global” approach toward a generalization of the idea of the cumulative set universe for fuzzy sets, there is also a recent more “local” one [Běhounek and Cintula, 200X] which only aims to give a unified treatment of a theory of fuzzy subsets of a given universe of discourse, i.e. which–in a suitable sense–restricts the considerations to the first level of the transfinite hierarchy (55).

The authors of [Běhounek and Cintula, 200X] use the (first-order) fuzzy logic ŁΠ, extended again with the $\triangle$-operator, as the background logical system. They take it as a two-sorted language with one sort of variables for objects of the universe of discourse and the other sort for fuzzy sets. The advantage of this choice is that (i) this logic is well understood, cf. e.g. [Horčik and Cintula, 2004], and that (ii) it has sufficiently high expressive power such that former approaches, like [Gottwald, 1993], which used a mixture of object and metalanguage considerations, can be unified and given in a uniform way within the language of ŁΠ, again with the primitive predicates $\in, =$. So one can e.g. express the fundamental *comprehension axiom* by the schema

$$\exists X \triangle \forall x (x \in X \leftrightarrow \varphi(x))$$

which has $\varphi(x)$ as an arbitrary formula of the language (not containing the set variable X free). And one can express the *axiom of extensionality* by

$$\forall x \triangle (x \in X \leftrightarrow x \in Y) \rightarrow X = Y\,.$$

---

[14] Here, as usual, by ${}^B A$ one denotes — for crisp sets A, B — the class of all functions from B into A.

This allows to denote fuzzy sets by class terms as in [Gottwald, 1993], with

(56) $a \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(a)$

as the guiding principle.

To guarantee the existence of fuzzy sets which are not crisp ones, one may either start with the logic $\text{Ł}\Pi\frac{1}{2}$ or add a specific *axiom of fuzziness* reading

$$\exists X \exists x (x \in X \leftrightarrow \neg_{\mathsf{L}} (x \in X)) \, .$$

Some examples shall illustrate the expressive power of this language:

$\{x \mid \neg_{\Pi}\neg_{\mathsf{L}}(x \in X)\}$	defines	the kernel of X ,
$\{x \mid \neg_{\Pi}\neg_{\Pi}(x \in X)\}$	defines	the support of X ,
$\{x \mid \triangle(\overline{\alpha} \to x \in X)\}$	defines	the (closed) α-cut of X ,

of course with $\overline{\alpha}$ as truth degree constant to denote the truth degree α.

A further expansion of the language with additional sorts of variables allows the authors to develop a machinery to discuss also fuzzy sets of higher level, and finally also a kind of fuzzy type theory and higher order fuzzy logics. Actually this is work in progress, partly contained in [Cintual, 2005].

14.2 Nonmonotonic fuzzy reasoning

One of the core areas for the application of logic in computer science is artificial intelligence. And inside AI, nonmonotonic reasoning has a prominent position.

So it is natural to ask how the basic ideas of nonmonotonic inference can be generalized from the crisp case to the fuzzy case, i.e. to the case in which either the knowledge comes e.g. with degrees of vagueness, or of confidence, or in which e.g. the defaults are accepted only to some degrees.

A first idea was offered in [Lehmke and Thiele, 1996]. This paper generalizes the circumscription approach in a straightforward way from classical logic to the infinite-valued Łukasiewicz logic L, and gives some basic properties of the nonmonotonic inference operator defined via minimal models.

Interesting new ideas, based upon the abstract approach toward fuzzy logic discussed in §13, have quite recently been offered in [Booth and Richter, 2005; Richter, 2003].

It is possible to define, quite similar to the standard case (6), for abstract fuzzy semantics $\mathcal{M}$ the *model class of* a fuzzy set Σ of formulas as

(57) $\mathbf{mod}_{\mathcal{M}}(\Sigma) = \{M \in \mathcal{M} \mid M \models_{\mathcal{M}} \Sigma\} \, ,$

and to define the *theory of* a class $\mathbb{K} \subseteq \mathcal{M}$ of models as

(58) $\mathbf{th}(\mathbb{K}) = \bigcup \{u \in \mathcal{F}_L(\mathbb{F}) \mid M \models u \text{ for all } M \in \mathbb{K}\} \, ,$

which means, in accordance with (53), that one has

$$\mathbf{th}(\mathbb{K}) = \bigcup\{u \in \mathcal{F}_L(\mathbb{F}) \mid u \subseteq \bigcap \mathbb{K}\} = \bigcap \mathbb{K}\,.$$

It is a routine matter to prove that for each fuzzy set $\Sigma \in \mathcal{F}_L(\mathbb{F})$ of formulas one obtains

$$\mathcal{C}^{\mathsf{sem}}(\Sigma) = \mathbf{th}(\mathbf{mod}_{\mathcal{M}}(\Sigma))\,,$$

i.e. $\mathbf{th}(\mathbf{mod}_{\mathcal{M}}(\Sigma))$ is a $\mathcal{C}^{\mathsf{sem}}$-theory.

Therefore it is possible to adapt within this abstract setting the model theoretic method of nonmonotonic inference which connects with each set Σ of formulas as its nonmonotonic inference hull $\mathbf{C}_{\sim}(\Sigma)$ the theory of a subclass $\Phi(\mathbf{mod}(\Sigma))$ of the class $\mathbf{mod}(\Sigma)$ of all models of Σ:

$$\mathbf{C}_{\sim}(\Sigma) = \mathbf{th}(\Phi(\mathbf{mod}(\Sigma)))\,,$$

e.g. the subclass of all normal or of all minimal models. In this generalized setting one can prove quite similar theoretical results as in the crisp case, as can be seen from [Richter, 2002].

Also another tool from nonmonotonic reasoning has a natural generalization to a fuzzy setting: *Poole systems* as introduced in [Poole, 1989]. Such a crisp Poole system P is determined by a pair (D, C) of sets of sentences understood as the relevant *defaults* and *constraints*. For each set Σ of formulas and a suitably chosen closure operator $\mathbf{C}$ it defines a class $\mathbf{E}_P$ of extensions by

$$\mathbf{E}_P(\Sigma) = \{\mathbf{C}(\Sigma \cup D_m) \mid D_m \subseteq D \text{ maximal w.r.t. consistency of } \Sigma \cup C \cup D_m\}\,,$$

and an inference operator $\mathcal{C}_P$ by

$$\mathcal{C}_P(\Sigma) = \bigcap \mathbf{E}_P(\Sigma)\,.$$

All these definitions allow, in the abstract setting of §13, a natural extension to the case of fuzzy sets of defaults and constraints. Details again may be found in [Richter, 2002; Richter, 2003].

However, even a more practical application becomes available: toward fuzzy belief revision.

A fuzzy belief base B is just a fuzzy set of formulas $B \in \mathcal{F}_L(\mathbb{F})$. The revision information (φ/a), understood as the fuzzy singleton of φ with membership degree a, tells that a "new" formula φ should be integrated with degree a. As in the AGM framework [Alchourron *et al.*, 1985] for the crisp case this may happen in the following steps, cf. [Booth and Richter, 2003; Booth and Richter, 2005]:

1. Form the family $B \perp (\varphi/a)$ of all maximal $X \in \mathcal{F}_L(B)$ consistent with (φ/a).
2. Select a subset $\gamma(B \perp (\varphi/a)) \subseteq B \perp (\varphi/a)$ and form its meet.

3. Add the revision information to get the revised belief base

$$B \star (\varphi/a) = \bigcap \gamma(B \perp (\varphi/a)) \cup (\varphi/a)\,.$$

The adaptation of this procedure to the case of the revision of fuzzy theories is not as straightforward as in the crisp case, but can also be handled sufficiently well with some extra care regarding the moment for taking (deductive) closures. Details are in [Richter, 2003].

BIBLIOGRAPHY

[Agliano *et al.*, 200X] P. Agliano, I. M. A. Ferreirim and F. Montagna. Basic hoops: an algebraic study of continuous t-norms. *Studia Logica* (200X, to appear).

[Alchourron *et al.*, 1985] C. Alchourron, P. Gärdenfors and D. Makinson. On the logic of theory change: partial meet contraction and revision functions, *J. Symbolic Logic* **50**, 510–530. 1985.

[Åqvist, 1962] L. Åqvist. Reflections on the logic of nonsense, *Theoria* **28**, 138–157, 1962.

[Avron, 1999] A. Avron. On the expressive power of three-valued and four-valued languages, *J. of Logic and Computation* **9**, 977–994, 1999.

[Baaz, 1996] M. Baaz. Infinite-valued Gödel logics with 0-1 projections and relativizations. In: *Gödel'96* (P. Hájek, ed.), Lecture Notes in Logic 6, Springer: New York, 23–33, 1996.

[Baaz *et al.*, 1993] M. Baaz, C. G. Fermüller, G. Quirchmayr and R. Zach. A many-valued logic for default reasoning, in: L. Gün/R. Onvural/E. Gelenbe (eds.), *Proc. 8th Internat. Symp. Computer and Information Sci.*, ISCIS VIII, Istanbul, Nov. 1993, Middle East Techn. Univ., Ankara, 476–479, 1993.

[Baudry, 1950] L. Baudry. *La Querelle des Future Contingents: Louvain 1465–1475*, Vrin, Paris, 1950.

[Běhounek and Cintula, 200X] L. Běhounek and P. Cintula. Fuzzy class theory, *Fuzzy Sets Systems* (to appear, 200X).

[Bell, 1985] J. L. Bell. *Boolean-Valued Models and Independence Proofs in Set Theory*. Oxford Logic Guides 12, Oxford University Press, Oxford, 1985.

[Belnap, 1977] N. D. Belnap. How a computer should think, in: G. Ryle (ed.) *Contemporary Aspects of Philosophy*. Oriel Press, Stockfield, 30–56, 1977.

[Belnap, 1977a] N. D. Belnap. A useful four-valued logic, in: J.M. Dunn/G. Epstein (eds.), *Modern Uses of Multiple-Valued Logic*. Reidel, Dordrecht, 8–37, 1977.

[Bernays, 1926] P. Bernays. Axiomatische Untersuchungen des Aussagenkalküls der "Principia Mathematica". *Math. Zeitschr.* **25**, 305–320, 1926.

[Beth, 1965] E. W. Beth *The Foundations of Mathematics*. North-Holland Publ. Comp., Amsterdam, 1965.

[Black, 1937] M. Black. Vagueness: an exercise logical analysis, *Philosophy of Science* **4**, 427–455, 1937.

[Black, 1963] M. Black. Reasoning with loose concepts, *Dialogue* **2**, 1–12, 1963.

[Blau, 1978] U. Blau. *Die dreiwertige Logik der Sprache: ihre Syntax, Semantik und Anwendung in der Sprachanalyse*. de Gruyter, Berlin, 1978.

[Bočvar, 1938] D. A. Bočvar. Ob odnom trechznačnom isčislenii i ego primenenii k analizu paradoksov klassičeskogo funkcional'nogo isčislenija. *Mat. Sbornik* **4** (46), 287–308, 1938.

[Bočvar, 1943] D. A. Bočvar. K voprosu o neprotivorečivosti odnogo trechznačnogo isčislenija. *Mat. Sbornik* **12** (54), 353–369, 1943.

[Bolc and Borowik, 1992] L. Bolc and P. Borowik. *Many-Valued Logics,* 1. Theoretical Foundations. Springer, Berlin, 1992.

[Booth and Richter, 2003] R. Booth and E. Richter. On revising fuzzy belief bases. In: *UAI 2003.* Proc. 19. Internat. Conf. on Uncertainty in Artificial Intelligence, Acapulco, Mexico (U. Kjaerulff, C. Meek eds.), Morgan Kaufmann: San Francisco, 81–88, 2003.

[Booth and Richter, 2005] R. Booth and E. Richter. On revising fuzzy belief bases, *Studia Logica* **80**, 29–61, 2005.

[Butnariu *et al.*, 1995] D. Butnariu, E. P. Klement and S. Zafrany. On triangular norm-based propositional fuzzy logics, *Fuzzy Sets Systems* **69**, 241–255, 1995.

[Burris and Sankappanavar, 1981] S. Burris and H. P. Sankappanavar. *A Course in Universal Algebra.* Springer: New York, 1981.

[Cat-Ho, 1973] Cat-Ho, Nguyen. *Generalized Post Algebras and Their Applications to Some Infinitary Many-Valued Logics.* Dissertationes Math., vol. 57, PWN, Warsaw, 1973.

[Chang, 1958] C. C. Chang. Algebraic analysis of many valued logics, *Transactions American Mathematical Society* **88**, 476–490, 1958.

[Cignoli, 1982] R. Cignoli. Proper n-valued Łukasiewicz algebras as $\mathfrak{S}$-algebras of Łukasiewicz n-valued propositional calculi, *Studia Logica* **41**, 3–16, 1982.

[Cignoli *et al.*, 2000] R. Cignoli, F. Esteva, L. Godo and A. Torrens. Basic Fuzzy Logic is the logic of continuous t-norms and their residua, *Soft Computing* **4**, 106–112, 2000.

[Cignoli *et al.*, 2000] R. Cignoli, I. D'Ottaviano and D. Mundici. *Algebraic Foundations of Many-Valued Reasoning.* Kluwer Acad. Publ., Dordrecht, 2000.

[Cintual, 2001] P. Cintula. The ŁΠ and ŁΠ$\frac{1}{2}$ propositional and predicate logics, *Fuzzy Sets Systems* **124**, 289–302, 2001.

[Cintula, 2001a] P. Cintula. An alternative approach to the ŁΠ logic, *Neural Network World* **124**, 561–575, 2001.

[Cintula, 2001b] P. Cintula. Advances in ŁΠ and ŁΠ$\frac{1}{2}$ logics, *Archive Math. Logic* **42**, 449–468, 2001.

[Cintual, 2005] P. Cintula. *From Fuzzy Logic to Fuzzy Mathematics.* Ph.D. Thesis, Technical University, Prague, 2005.

[Cleave, 1974] J. P. Cleave. The notion of logical consequence in the logic of inexact predicates, *Zeitschr. math. Logik Grundl. Math.* **20**, 307–324, 1974.

[Denecke and Wismath, 2002] K. Denecke and S. L. Wismath. *Universal Algebra and Applications in Theoretical Computer Science*, Chapman & Hall/CRC: Boca Raton/FL, 2002.

[di Nola *et al.*, 2002] A. di Nola, G. Georgescu and A. Iorgulescu. Pseudo-BL algebras. I and II, *J. Multiple-Valued Logic* **8**, 671–750, 2002.

[Dummett, 1959] M. Dummett. A propositional calculus with denumerable matrix, *J. Symbolic Logic* **24**, 97–106, 1959.

[Dunn, 1976] J. M. Dunn. Intuitive semantics for first-degree entailments and 'coupled trees', *Philosophical Studies* **29**, 149–168, 1976.

[Dunn, 1998] J. M. Dunn. A comparative study of various semantical treatments of negation: a history of formal negation, in: D. Gabbay/H. Wansing (eds.), *What is Negation.* Kluwer, Dordrecht, 23–51, 1988.

[Dwinger, 1968] Ph. Dwinger. Generalized Post algebras, *Bull. Acad. Polon. Sci.*, Sér. Sci. Math., Astronom., Phys., **16**, 559–563, 1968.

[Dwinger, 1977] Ph. Dwinger. A survey of the theory of Post algebras and their generalizations, in: J.M. Dunn/G. Epstein (eds.), *Modern Uses of Multiple-Valued Logic.* Reidel, Dordrecht, 53–75, 1977.

[Ebbinghaus, 1969] H.-D. Ebbinghaus. Über eine Prädikatenlogik mit partiell definierten Prädikaten und Funktionen, *Archiv math. Logik Grundlagenforsch.* **12**, 39–53, 1969.

[Epstein, 1960] G. Epstein. The lattice theory of Post algebras, *Transactions American Math. Soc.* **95**, 300–317, 1960. Reprinted in: D.C. Rine (ed.) *Computer Science and Multiple Valued Logic.* North-Holland Publ. Comp., Amsterdam 1977, 17–34.

[Epstein, 1990] G. Epstein. *The Semantic Foundations of Logic.* Vol.1: Propositional Logics. Kluwer, Dordrecht, 1990.

[Epstein and Rasiowa, 1990] G. Epstein and H. Rasiowa. Theory and uses of Post algebras of order $\omega + \omega^*$. I, in: *20th Internat. Symp. Multiple-Valued Logic, Charlotte/NC, 1990,* Proceedings, IEEE Computer Soc., New York, 42–47, 1990.

[Epstein and Rasiowa, 1991] G. Epstein and H. Rasiowa. Theory and uses of Post algebras of order $\omega + \omega^*$. Part II, in: *21st Internat. Symp. Multiple-Valued Logic, Victoria/B.C., 1991*, Proceedings, IEEE Computer Soc., New York, 248–254, 1991.

[Esteva and Godo, 1999] F. Esteva and L. Godo. Putting together Łukasiewicz and product logics, *Mathware and Soft Computing* **6**, 219–234, 1999.

[Esteva and Godo, 2001] F. Esteva and L. Godo. Monoidal t-norm based logic: towards a logic for left-continuous t-norms, *Fuzzy Sets Systems* **124**, 271–288, 2001.

[Esteva *et al.*, 2000] F. Esteva, L. Godo, P. Hájek and M. Navara. Residuated fuzzy logic with an involutive negation, *Archive Math. Logic* **39**, 103–124, 2000.

[Esteva *et al.*, 2003] F. Esteva, L. Godo, P. Hájek and F. Montagna. Hoops and fuzzy logic, *Journ. Logic and Comput.* **13**, 531–555, 2003.

[Esteva *et al.*, 2001] F. Esteva, L. Godo and F. Montagna. The ŁΠ and ŁΠ$\frac{1}{2}$ logics: two complete fuzzy systems joining Łukasiewicz and product logics, *Archive Math. Logic* **40**, 39–67, 2001.

[Esteva *et al.*, 2004] . Esteva, L. Godo and F. Montagna. Equational characterization of the subvarieties of BL generated by t-norm algebras, *Studia Logica* **76**, 161–200, 2004.

[Finn, 1971] V. K. Finn. Ob aksiomatizacii nekotorych trechznačnych logik, *Naučno-Techničeskaja Informacija*, ser. 2, no. 11, 16–20, 1971.

[Finn, 1974] V. K. Finn. A criterion of functional completeness for B3, *Studia Logica* **33**, 121–125, 1974.

[Finn, 1974a] V. K. Finn. Aksiomatizacii nekotorych trechznačnych isčislenij vyskazyvanij i ich algebr, in: *Filosofija v Sovremennom Mire. Filosofija i Logika*. Nauka, Moscow, 398–438, 1974.

[Fisch and Turquette, 1966] M. Fisch and A. R. Turquette. Peirce's triadic logic. *Trans. Charles Sanders Peirce Society* **2**, 71–85, 1966.

[Fitting, 1988] M. C. Fitting. Logic programming on a topological bilattice, *Fundamenta Informaticae* **11**, 209–218, 1988.

[Fitting, 1989] M. C. Fitting. Bilattices and the theory of truth, *J. Philosophical Logic* **18**, 225–256, 1989.

[Flondor *et al.*, 2001] P. Flondor, G. Georgescu and A. Iorgulescu. Pseudo t-norms and pseudo-BL algebras, *Soft Computing* **5**, 355–371, 2001.

[Gerla, 2001] G. Gerla. *Fuzzy Logic.* Mathematical Tools for Approximate Reasoning, Trends in Logic, vol. 11, Kluwer Acad. Publ., Dordrecht, 2001.

[Giles, 1976] R. Giles. Łukasiewicz logic and fuzzy set theory, *Internat. J. Man-Machine Studies* **8**, 313–327, 1976.

[Ginsberg, 1987] M. L. Ginsberg. Multi-valued logics, in: M.L. Ginsberg (ed.), *Readings in Nonmonotonic Reasoning*. Morgan Kaufman, Los Altos/Ca., 251–255, 1987.

[Ginsberg, 1988] M. L. Ginsberg. Multivalued logic: a uniform approach to reasoning in artificial intelligence, *Computational Intelligence* **4**, 265–315, 1988.

[Goddard and Routley, 1973] L. Goddard and R. Routley. *The Logic of Significance and Context.* Scottish Academic Press, Edinburgh, 1973.

[Gödel, 1932] K. Gödel. Zum intuitionistischen Aussagenkalkül, *Anzeiger Akademie der Wissenschaften Wien*, Math.-naturwiss. Klasse **69**, 65–66, 1932; also: *Ergebnisse eines mathematischen Kolloquiums* H. 4 (1933) 40.

[Goguen, 1968] J. A. Goguen. The logic of inexact concepts, *Synthese* **19** 325–373, 1968–69.

[Gottwald, 1984] S. Gottwald. Fuzzy set theory. Some aspects of the early development. In: *Aspects of Vagueness* (H.-J. Skala, S. Termini, E. Trillas eds.), Theory and Decision Libr., vol. 39, Reidel, Dordrecht, 13–29, 1984.

[Gottwald, 1993] S. Gottwald. *Fuzzy Sets and Fuzzy Logic.* The Foundations of Application — From a Mathematical Point of View, Vieweg, Braunschweig/Wiesbaden and Teknea, Toulouse, 1993.

[Gottwald, 2001] S. Gottwald. *A Treatise on Many-Valued Logics.* Studies in Logic and Computation, vol. 9, Research Studies Press, Baldock, 2001.

[Gottwald and Jenei, 2001] S. Gottwald and S. Jenei. A new axiomatization for involutive monoidal t-norm based logic, *Fuzzy Sets Systems* **124**, 303–307, 2001.

[Gottwald and Hájek, 2005] S. Gottwald and P. Hájek. T-norm based mathematical fuzzy logics. In: *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms* (E.P. Klement and R. Mesiar, eds.), Elsevier, Dordrecht, 275–299, 2005.

[Grätzer, 1979] G. Grätzer. *Universal Algebra.* Springer, Berlin, 1979.

[Grigolia, 1977] R. Š. Grigolia. Algebraic analysis of Łukasiewicz–Tarski's n-valued logical systems, in: R. Wójcicki/G. Malinowski (eds.) *Selected Papers on Łukasiewicz Sentential Calculi.* Ossolineum, Wroclaw, 81–92, 1977.

[Gupta and Belnap, 1993] A. Gupta and N. Belnap. *The Revision Theory of Truth.* MIT Press, Cambridge/Mass, 1993.

[Hájek, 1998] P. Hájek. Basic fuzzy logic and BL-algebras, *Soft Computing* **2**, 124–128, 1998.

[Hájek, 1998] P. Hájek. *Metamathematics of Fuzzy Logic.* Trends in Logic, vol. 4, Kluwer Acad. Publ., Dordrecht, 1998.

[Hájek, 2003] P. Hájek. Embedding standard BL-algebras into non-commutative pseudo-BL-algebras, *Tatra Mt. Math. Publ.* **27**, 125–130, 2003.

[Hájek, 2003a] P. Hájek. Fuzzy logics with non-commutative conjunctions, *Journ. Logic and Comput.* **13**, 469–479, 2003.

[Hájek, 2003b] P. Hájek. Observations on non-commutative fuzzy logics, *Soft Computing* **8**, 28–43, 2003.

[Hájek, 2005] P. Hájek. Fleas and fuzzy logic, *J. Multiple-Valued Logic and Soft Computing* **11**, 137–152, 2005.

[Hájek *et al.*, 1996] P. Hájek, L. Godo and F. Esteva. A complete many-valued logic with product-conjunction, *Arch. Math. Logic* **35**, 191–208, 1006.

[Hájek and Haniková, 2003] P, Hájek and Z. Haniková. A development of set theory in fuzzy logic. In: *Beyond Two.* Theory and Applications of Multiple-Valued Logic (M. Fitting, E. Orlowska eds.), Physica-Verlag, Heidelberg, 273–285, 2003.

[Hájek *et al.*, 2000] P. Hájek, J. Paris and J. Shepherdson. Rational Pavelka predicate logic is a conservative extension of Łukasiewicz predicate logic, *J. Symbolic Logic* **65**, 669–682, 2000.

[Halldén, 1949] S. Halldén. *The Logic of Nonsense.* Uppsala Univ., Uppsala, 1949.

[Haniková, 2001] Z. Haniková. Standard algebras for fuzzy propositional calculi, *Fuzzy Sets Systems* **124**, 309–320, 2001.

[Haniková, 2003] Z. Haniková. *Mathematical and Metamathematical Properties of Fuzzy Logic*, Ph.D. Thesis, Charles University Prague, 2003.

[Hekrdla *et al.*, 2003] J. Hekrdla, E. P. Klement and M. Navara. Two approaches to fuzzy propositional logics, *J. Multiple-Valued Logic and Soft Computing* **9**, 343–360, 2003.

[Horčik and Cintula, 2004] R. Horčik and P. Cintula. Product Łukasiewicz logic, *Archive Math. Logic* **43**, 477–503, 2004.

[Horn, 1989] L. R. Horn. *A Natural History of Negation.* University of Chicago Press, Chicago, 1989.

[Jaśkowski, 1936] S. Jaśkowski. Recherches sur le système de la logique intuitioniste. In: *Actes du Congrès Internat. de Philosophie Scientifique 1936*, vol. 6, Actualitès Scientifique et Industriel, vol. 393, Hermann, Paris, 58–61, 1936.

[Jenei and Montagna, 2002] S. Jenei and F. Montagna. A proof of standard completeness for Esteva and Godo's logic MTL, *Studia Logica* **70**, 183–192, 2002.

[Jenei and Montagna, 2003] S. Jenei and F. Montagna. A proof of standard completeness for non-commutative monoidal t-norm logic. *Neural Network World* **13**, 481–488, 2003.

[Kearns, 1974] J. T. Kearns. Vagueness and failing sentences, *Logique et Analyse*, N. S. **17**, 301–315, 1974.

[Kearns, 1979] J. T. Kearns. The strong completeness of a system for Kleene's three-valued logic, *Zeitschr. math. Logik Grundl. Math.* **25**, 61–68, 1979.

[Klaua, 1966] D. Klaua. Über einen zweiten Ansatz zur mehrwertigen Mengenlehre, *Monatsberichte Deut. Akad. Wissensch. Berlin* **8**, 161–177, 1966.

[Kleene, 1938] S. C. Kleene. On notation for ordinal numbers. *J. Symbolic Logic* **3**, 150–155, 1938.

[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics,* North-Holland Publ. Comp.,Amsterdam and van Nostrand, New York, 1952.

[Klement *et al.*, 2000] E. P. Klement, R. Mesiar and E. Pap. *Triangular Norms.* Kluwer Acad. Publ., Dordrecht, 2000.

[Körner, 1966] S. Körner. *Experience and Theory.* Routledge & Kegan Paul, London, 1966.

[Kripke, 1975] S. A. Kripke. Outline of a theory of truth, *J. of Philosophy* **72**, 690–716, 1975.

[Kühr, 2003] J. Kühr. Pseudo-BL algebras and PRl-monoids. *Mathematica Bohemica* 128 199–208, 2003.

[Lehmke and Thiele, 1996] S. Lehmke and H. Thiele. On fuzzy circumscription. In: *EUSFLAT '96.* Proc. Fourth Europ. Congr. Intell. Techniques and Soft Computing, Aachen 1996, vol.1, Aachen, 641–645, 1996.

[Lewis and Langford, 1932] C. J. Lewis and C. H. Langford. *Symbolic Logic,* Dover, New York, 1932.

[Lovett, 1900] E. O. Lovett. Mathematics at the International Congress of Philosophy, Paris 1900. *Bull. Amer. Math. Soc.* **7**, 157–183, 1900–01.

[Łukasiewicz, 1913] J. Łukasiewicz. *Die logischen Grundlagen der Wahrscheinlichkeits- rechnung*, Krakow [cf. also his (1970)], 1913.

[Łukasiewicz, 1920] J. Łukasiewicz. O logice trójwartościowej. *Ruch Filozoficzny* **5**, 170–171, 1920.

[Łukasiewicz, 1930] J. Łukasiewicz. Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls, *Comptes Rendus Séances Société des Sciences et Lettres Varsovie*, cl. III, **23**, 51–77, 1930.

[Łukasiewicz, 1935] J. Łukasiewicz. Zur Geschichte der Aussagenlogik. *Erkenntnis* **5**, 111–131, 1935.

[Łukasiewicz, 1953] J. Łukasiewicz. A system of modal logic, *J. Computing Systems* **1**, 111–149, 1953.

[Łukasiewicz, 1970] J. Łukasiewicz. *Selected Works* (L. Borkowski, ed.). North-Holland Publ. Comp., Amsterdam, 1970.

[Łukasiewicz and Tarski, 1930] J. Łukasiewicz and A. Tarski. Untersuchungen über den Aussagenkalkül. *Comptes Rendus Séances Société des Sciences et Lettres Varsovie*, cl. III, **23**, 30–50, 1930.

[Martin and Woodruff, 1975] R. L. Martin and P. W. Woodruff. On representing 'True-in-L' in L, *Philosophia* **5**, 213–217, 1975.

[McColl, 1897] H. McColl. Symbolic reasoning. II, *Mind*, N.S., **6**, 493–510, 1897.

[McGee, 1991] V. McGee. *Truth, Vagueness, and Paradox.* Hackett Publ. Comp., Indianapolis/Ind, 1991.

[Michalski, 1937] K. Michalski. Le problème de la volonté à Oxford et à Paris au XIVe siècle. *Studia Philosophica* **2**, 233–365, 1937.

[Moisil, 1940] G. C. Moisil. Recherches sur les logiques non chrysipiennes, *Annales Scientifiques de l'Université de Jassy* **26**, 431–466, 1940.

[Moisil, 1941] G. C Moisil. Notes sur les logiques non-chrysipiennes, *Annales Scientifiques de l'Université de Jassy* **27**, 86–98, 1941.

[Mostowski, 1957] A. Mostowski. On a generalization of quantifiers. *Fund. Math.* **44**, 12–36, 1957.

[Novák *et al.*, 1999] V. Novák, I. Perfilieva and J. Močkoř. *Mathematical Principles of Fuzzy Logic.* Kluwer, Boston, 1999.

[Patzig, 1973] G. Patzig. Aristotle, Łukasiewicz and the origins of many-valued logic. In: *Logic, Methodology and Philosophy of Science IV* (P. Suppes et al. eds.), PWN, Warsaw and North-Holland Publ. Comp., Amsterdam, 921–929, 1973.

[Pavelka, 1979] J. Pavelka. On fuzzy logic. I–III, *Zeitschr. math. Logik Grundl. Math.* **25**, 45–52; 119–134; 447–464, 1979.

[Peirce, 1931] C. S. Peirce. *Collected Papers* (C. Hartshorne et al. eds.), Harvard Univ. Press, Cambridge/MA, 1931–58.

[Piróg-Rzepecka, 1977] K. Piróg-Rzepecka. *Systemy nonsense-logics.* PWN, Warsaw–Wroclaw, 1977.

[Poole, 1989] D. Poole. What the lottery paradox tells us about default reasoning. In: *KR '89: Proc. First Internat. Conf. Principles of Knowl. Repres. and Reas.* (H.J. Levesque, R.J. Brachman, R. Reiter eds.), Morgan Kaufmann: San Matteo/Ca., 333–340, 19889.

[Post, 1921] E. L. Post. Introduction to a general theory of elementary propositions. *Amer. J. Math.* **43**, 163–185, 1921.

[Priest and Sylvan, 1992] G. Priest and R. Sylvan. Simplified semantics for basic relevant logics, *J. Philosophical Logic* **21**, 217–232, 1992.

[Pynko, 1999] A. P. Pynko. Functional completeness and axiomatizability within Belnap's four-valued logic and its expansions, *J. Applied Non-Classical Logics* **9**, 61–106, 1999.

[Raju, 1954] P. T. Raju. The principle of four-cornered negation in Indian philosophy, *Review of Metaphysics* **7**, 694–713, 1954.

[Rasiowa, 1973] H. Rasiowa. On generalised Post algebras of order ω^+ and ω^+-valued predicate calculi, *Bull. Acad. Polon. Sci.*, Sér. Sci. Math., Astronom., Phys., **21**, 209–219, 1973.

[Rescher, 1969] N. Rescher. *Many-Valued Logic,* McGraw-Hill, New York, 1969.

[Restall, 1995] G. Restall. Four-valued semantics for relevant logics (and some of their rivals), *J. Philosophical Logic* **24**, 139–160, 1995.

[Richter, 2002] E. Richter. Nonmonotonic inference operators for fuzzy logic. In: *Proc. 9th Internat. Workshop on Nonmonotonic Reasoning, NMR'2002, Toulouse, France* (S. Benferhat, E. Giunchiglia, eds.), Toulouse, 321–329, 2002.

[Richter, 2003] E. Richter. *Nonmonotonic Fuzzy Inference.* Ph.D. Thesis, Leipzig University, 2003.

[Rine, 1977] D. C. Rine, ed. *Computer Science and Multiple Valued Logic.* North-Holland Publ. Comp., Amsterdam, 1977. 2nd rev. ed. 1984.

[Rosenbloom, 1942] P. C. Rosenbloom. Post algebras. I. Postulates and general theory, *American J. Mathematics* **64**, 167–188, 1942.

[Rosser and Turquette, 1952] J. B. Rosser and A. R. Turquette. *Many-Valued Logics*, North-Holland Publ. Comp., Amsterdam, 1952.

[Segerberg, 1965] K. Segerberg. A contribution to nonsense-logics, *Theoria* **31**, 199–217, 1965.

[Šestakov, 1964] V. I. Šestakov. O vzaimootnošenii nekotorych trechznačnych logikach isčislenij, *Uspechi Matematičeskich Nauk* **19**, 177–181, 1964.

[Sette, 1973] A. M. Sette. On the propositional calculus P^1, *Mathematica Japonica* **18**, 173–180, 1973.

[Sette and Carnielli, 1995] A. M. Sette and W. A. Carnielli. Maximal weakly-intuitionistic logics, *Studia Logica* **55**, 181–203, 1995.

[Słupecki, 1936] J. Słupecki. Der volle dreiwertige Aussagenkalkül. *C. R. Séances Soc. Sci. et Lettr. Varsovie,* cl. III, **29**, 9–11, 1936.

[Takeuti and Titani, 1984] G. Takeuti and S. Titani. Intuitionistic fuzzy logic and intuitionistic fuzzy set theory, *J. Symbolic Logic* **49**, 851–866, 1984.

[Takeuti and Titani, 1992] G. Takeuti and S. Titani. Fuzzy logic and fuzzy set theory, *Archive Math. Logic* **32**, 1–32, 1992.

[Tarski, 1930] A. Tarski. Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften, *Monatshefte Mathematik Physik* **37**, 361–404, 1930.

[Traczyk, 1967] T. Traczyk. On Post algebras with uncountable chain of constants. Algebras and homomorphisms, *Bull. Acad. Polon. Sci.*, Sér. Sci. Math., Astronom., Phys., **15**, 673–680, 1967.

[Turquette, 1967] A. R. Turquette. Peirce's Phi and Psi operators for triadic logic. *Trans. Charles Sanders Peirce Society* **3**, 66–73, 1967.

[Turunen, 1995] E. Turunen. Well-defined fuzzy sentential logic, *Math. Logic Quarterly* **41**, 236–248, 1995.

[Visser, 1984] A. Visser. Four-valued semantics and the Liar, *J. Philosophical Logic* **13**, 181–217, 1984.

[Wajsberg, 1977] M. Wajsberg. . Axiomatization of the three-valued propositional calculus. In: *M. Wajsberg: Logical Works* (S.S. Surma ed.), Ossolineum, Wroclaw, 12–29, 1977.

[Wójcicki, 1988] R. Wójcicki. *Theory of Logical Calculi: Basic Theory of Consequence Operations*, Kluwer Acad. Publ., Dordrecht, 1988.

[Zadeh, 1965] L. A. Zadeh. Fuzzy sets, *Information and Control* **8**, 338–353, 1965.

RELEVANCE LOGICS

Katalin Bimbó

Relevance logics emerged in the middle of the 20th century as a characterization of *implication* and *entailment.* From early on it was known that classical logic validates certain weird inferences and assigns the truth-value 'true' to conditional statements of a peculiar kind. It was "understood" that classically there need not be a connection between the premises and the conclusion or between the antecedent and the consequent of a conditional — except certain patterns of truth values.

Of course, one could try to shrug away the irrelevance of the classical consequence relation by saying that it is a "pragmatic issue." However, it is not. Relevance logics show that one of the causes of the problem can be pinpointed as an axiom (in a frequently used formulation) of classical logic, and *relevant inferences* can be captured *mathematically precisely* — by logical calculi. Accordingly, we use the term 'relevance logics' primarily to refer to logics that do not contain $\mathcal{A} \to (\mathcal{B} \to \mathcal{A})$ as an axiom (or a theorem).[1] As a word of caution, however, we note that once the overpowering dominance of classical logic has been successfully challenged (and it has been), it becomes clear that there is no single superior way to formalize what is a logically valid inference. There are several ways, though some are more plausible and useful than others.[2]

One of the first relevance logics is Church's *weak implicational calculus* (introduced in Church [1951] and later denoted as $\mathbf{R}_{\to}$), which is axiomatized by the principal type schemes of combinators that can define $\lambda\mathsf{I}$-abstraction. Indeed, Church defined his calculus by replacing $\mathcal{A} \to (\mathcal{B} \to \mathcal{A})$ (the principal type scheme of K) by $\mathcal{A} \to \mathcal{A}$ (the principal type scheme of I) in the positive implicational logic.[3] Ackermann's and further on, Anderson and Belnap's aim was more inherently philosophical: to formalize the notion of strong implication and of entailment

---

[1] There are other implicational theorems that turn a relevance logic into a "somewhat relevant logic," e.g., the "mingle axiom" $\mathcal{A} \to (\mathcal{A} \to \mathcal{A})$, that is closely related to K. There are "classical relevance logics," that include "Boolean negation" (hence permit certain irrelevance) but conservative over their "nonclassical" fragment.

[2] Classical logic is, of course, familiar to many people due to its prevalence in *curricula.* (See e.g., Enderton [2001] — a standard introduction in logic.) However, nonclassical logics appeared as early as (if not earlier than) the second decade of the 20th century, in the form of modal logics, many-valued logics, intuitionistic logic etc. A historical survey of relevance logics is Restall [200+]. We include historical figures and facts only for purposes of illustration and to fix some points for later reference.

[3] *Combinatory logic* was invented (independently) by Schönfinkel and by Curry; the λ-*calculi* are due to Church. We assume a certain familiarity with these logics — those aspects that are needed for our purposes here are introduced in Hindley [1997]. Incidentally, we note that the role of the type of K in the implicational fragment of classical logic is highlighted when that is formalized by adding K to Peirce's law and suffixing. By a result of J. Łukasiewicz, $\mathcal{A} \to (\mathcal{B} \to \mathcal{A})$

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

(see Ackermann [1956], Anderson and Belnap [1962]). Their research direction produced a wealth of mathematically and philosophically interesting results about some of the *best known relevance logics* (such as **T**, **E** and **R**), as well as their fragments, extensions, etc. Similar ideas emerged in other areas, for instance, in computer science. One of the relevance logics most recently introduced (that is also supported with an abundance of intuitive motivation) is *linear logic* **L**, which in effect adds a modality to one of the earlier relevance logics.[4] Thereby, linear logic hits a middle ground between allowing and disallowing irrelevant inferences — by keeping track of them. The connections to mathematics, philosophy and computer science that we mentioned above are just a small sample of the origins and influences of relevance logics. Other disciplines including informatics, artificial intelligence, linguistics, cognitive science (to mention some) are suitable areas to use and utilize relevance logics and, perhaps, some other logics too that are *nonmonotonic* (that is, they reject arbitrary additions of new premises) or that adhere to a *variable sharing* property (some — e.g., the Lambek calculi — to an even stricter one than relevance logics).

A chapter cannot cover — probably, could not even list — all the relevance logics, let alone can it state all the theorems, metatheorems, lemmas or give their proofs. The result of an attempt to squeeze everything about relevance logics into one chapter would not only be foolish but most likely result in something quite useless. Furthermore, there are comprehensive volumes and several recent survey articles on the topic. There are two two-volume monographs: Anderson and Belnap [1975], Anderson, Belnap and Dunn [1992] and Routley, Meyer, Plumwood and Brady [1982], Brady [2003] — each being the result of collaborative efforts of several logicians over quite a few years. (Anderson *et al.* [1992] also contains a bibliography.) A recent book that is primarily about the philosophical underpinnings of relevance logics is Mares [2004]. Handbook chapters and survey articles include Dunn [1986] (that has been updated as Dunn and Restall [2002]), as well as Mares and Meyer [2001] and Mares [2002].

Our aim here is to introduce and illustrate some of the motivating ideas that lead to the formal systems called relevance logics. We also hope to give a flavor of the excitement that arises from the technical details of some proofs and from the mathematical subtleties of certain results.[5] We chose as our central example one of the preeminent relevance logics: *the logic of ticket entailment.* The most recent surveys concentrated on **R** (and to a lesser extent on **E**), thus, it seems reasonable to bring into the foreground another well-motivated relevance logic. (Incidentally,

---

may then be replaced by any theorem of the form $\mathcal{C} \to (\mathcal{A} \to \mathcal{B})$, including the case when $\mathcal{A} \to \mathcal{B}$ is a theorem itself. Replacing $\mathcal{A} \to (\mathcal{B} \to \mathcal{A})$ with $\mathcal{A} \to \mathcal{A}$, however, yields only a proper sublogic of classical logic.

[4] Girard [1987] seems to be the first (English) exposition of linear logic.

[5] The main ideas and research approaches in relevance logics as well as the bulk of the results mentioned here come from published articles and books by others. Needless to say that I made an effort to report the results of the original works faithfully and to provide accurate references everywhere. Nonetheless, some lemmas, theorems etc. included here are new (as far as I know), hence, no references to other publications are given.

sometimes **T** is technically more challenging than **R**.) In the last section we whip through an array of logics and results — though we put a special emphasis on four well-known systems — to round out the picture.

1 A PROBLEM WITH CLASSICAL "IMPLICATION"

The propositional fragment of classical logic has an *undeniable splendor* stemming from its simple two-valued semantics. Classical logic is *Post complete*, that is, it forms an "upper limit" of the set of potential theorems — the inclusion of any new theorems makes classical logic inconsistent.[6] However, a discrepancy between inferences validated by classical logic and inferences considered acceptable was known from the very beginning. Similarly, it was realized long ago that there is a wide gap between true conditionals and "if…, then…" sentences which are truthfully stated.

Frege writes in his *Begriffsschrift*:

> For example, let $\vdash A$ stand for $3 \times 7 = 21$ and B for the circumstance that the sun is shining. … There need not exist a causal connection between the two contents.

Then, two paragraphs later he states:

> The causal connection inherent in the word "if", however, is not expressed by our signs, even though only such a connection can provide ground for a judgement of the kind under consideration. Frege [1967, 14].

Hilbert and Ackermann practically start their textbook — which was widely translated and used for many years — with a disclaimer:

> The compound sentence "if X then Y" is not to be taken to indicate a relation of premise to conclusion or of cause and effect. Rather, the sentence $X \to Y$ is always true if X is a false and also if Y is a true sentence. Hilbert and Ackermann [1950, 4].

One of their examples of a true "if…, then…" statement is: "If 2 times 2 equals 5, then snow is black."

Of course, we are not belittling the importance of Frege's attempt to sift out from the so-called "traditional logic" various psychological, metaphysical and linguistic factors, which were mixed up with accounts of valid patterns of inference.[7]

---

[6] Of course, this does not mean that a logic cannot include theorems that look like non-theorems of classical logic (when connectives are "matched"). Indeed, there are such logics, but they omit some of the theorems, connectives, etc. of classical logic.

[7] Incidentally, the only causal connections that we are to be concerned about are the *logical ones.*

Neither are we denouncing Hilbert and Ackermann, two well-known and accomplished logicians of the past century, though they surely must have been aware of the connection between implications and deductions in classical logic when they wrote the statements cited above (cf. especially 23–27). Rather, we quoted their writings to show that they knew that classical logic does not formalize *actual reasoning*, which is, perhaps, a minor deficiency. It is obvious that classical logic does not model everyday reasoning, if for no other reason because it is not wholly unparallelled that people misstep in reasoning.[8] We contend that classical logic does not formalize *idealized reasoning* either (that is, reasoning when we disregard mishaps, mistakes, errors and such), because classical logic validates some unusual and odd inferences.[9]

Some might opine that classical logic, perhaps, models *mathematical reasoning* — eventually, Boole and Ackermann, and especially Hilbert, are well-known as mathematicians (not as philosophers, linguists or psychologists). An imagined situation showing classical logic deficient as a tool in mathematical reasoning is described in Anderson and Belnap [1975, §3]. We intend to advance an argument leading to the same conclusion based on *actual happenings.* A somewhat forgotten fact is that Cantor knew (toward the end of the 1890s) that the set theory he invented was inconsistent (cf. Zermelo [1962, 443–451]). His argument was based on his proof that the cardinality of a powerset is always strictly greater than the cardinality of the set one started with, together with allowing any collection to be a set. Cantor used various terms for "collections," however, the realization of this inconsistency led him to label certain multiplicities inconsistent (or absolutely infinite). (In the Bernays–von Neumann–Gödel (BNG) axiomatization of set theory these "collections" turn out to be *proper classes.*) A more widely known fact is that he not only invented but also tried to prove the continuum hypothesis (CH) with no success. (From the work of Gödel [1938] and Cohen [1963; 1964], of course, it is clear why Cantor could not prove GCH (or CH).) What is interesting for our purposes here is to note that there is absolutely no record (as far as I know) to indicate or allege that Cantor attempted to infer CH from the existence of "inconsistent collections." That is, he did *not reason* according to classical logic, because he did not infer from the contradiction he discovered (which was later named "Cantor's paradox") the truth of the conjecture he made (the CH). Perhaps, this is the sort of situation which was meant by Wittgenstein when he claimed that inconsistency in mathematics (unlike in everyday life) is harmless, because no conclusion should be drawn from a (mathematical) contradiction. (Cf. Diamond [1976, 209, 213–220].)

There are at least two ways to look at our brief chronicle on Cantor's paradox as far as the pros and cons for classical logic are in question. The discovery of a contradiction led Cantor (and then others) to seek ways to avoid that particular contradiction — by marking certain kinds of collections as "different" (as in BNG) or by totally excluding them (as in ZF(C), the Zermelo–Fraenkel axiomatization of set theory (with the axiom of choice (AC))). This approach is in accordance with what we said about classical logic being Post complete. On the other hand, no sane mathematician tries to draw arbitrary conclusions from contradictions. The only "meta-conclusion" that (sometimes) seems to be drawn from a contradiction is that that piece of mathematics is badly flawed. Therefore, a calculus of implication (better than what classical logic provides) should also preclude the conclusion of an arbitrary proposition from a contradiction. This aspect of *mathematical reasoning* — together with others that we overview in §3 — is faithfully modeled by relevance logics.

---

[8] There is a vast amount of literature exhibiting that $\supset$ does not model natural language conditionals or counterfactuals either. It seems to us unnecessary to elaborate here on this (quite obvious) claim beyond noting the fact.

[9] There are famous and quite amusing examples (involving guinea pigs) that show the oddities of classical logic. Such sentences are given, for example, by Dunn [1986, 123] as well as in Mares [2004, §7.2].

Classically not only the inference from $\mathcal{A} \wedge \neg\mathcal{A}$ to $\mathcal{B}$ is correct, in customary notation, $(1')$ $\mathcal{A} \wedge \neg\mathcal{A} \vdash \mathcal{B}$, but also the conditional (1) $(\mathcal{A} \wedge \neg\mathcal{A}) \supset \mathcal{B}$ is valid. Notice that because of the rule *modus ponens* and of the *deduction theorem* it is justified to talk about inference patterns (with finitely many premises) and conditionals in tandem (as we did above), or to talk about just one of them in lieu of both. As it happens, (1) and $(1')$ are not the only problematic conditional-cum-inference pattern of classical logic. (1) is one of the so-called "paradoxes of material implication."[10] (1)–(6) are valid formulas of classical logic, which are counterintuitive when the connective $\supset$ is interpreted as implication. (The implausibility of the interpretation can be reinforced by choosing suitably disparate sentences in place of $\mathcal{A}$ and $\mathcal{B}$.)

(1) $(\mathcal{A} \wedge \neg\mathcal{A}) \supset \mathcal{B}$ — ex falso quodlibet,
(2) $\mathcal{A} \supset (\mathcal{B} \vee \neg\mathcal{B})$ — verum ex quodlibet,
(3) $\mathcal{A} \supset (\mathcal{B} \supset \mathcal{A})$ — positive paradox (K),
(4) $\mathcal{A} \supset (\mathcal{B} \supset \mathcal{B})$ — negative paradox (KI),
(5) $\mathcal{A} \supset (\neg\mathcal{A} \supset \mathcal{B})$ — ex falso quodlibet$_2$,
(6) $(\mathcal{A} \supset \mathcal{B}) \vee (\mathcal{B} \supset \mathcal{A})$ — linearity.[11]

Under the "$\supset$ is implication" interpretation (1) and (5) — in two different forms — state that a contradiction implies anything, while (2) states the dual. The positive paradox says that a sentence implies that any sentence implies it. The negative paradox can be read as: any formula implies an identity. Perhaps, one of the most strange classically valid formulas is (6), which says that given two arbitrary formulas (at least) one implies the other. This formula is similar to that expressing connectedness of a (weak) partial order. Of course, because of the two valued interpretations for classical logic, the validity of this formula is no surprise — **2** (the smallest nontrivial Boolean algebra) is, indeed, linearly ordered.

---

[10] 'Paradox' appears as a label for very different sorts of phenomena. Cantor's, Russell's, Curry's paradoxes each show a theory — naive or formalized — to be inconsistent or trivial. The so-called "paradoxes of material implication" are more similar to Skolem's paradox; these refer to discrepancies between intuitions and what certain formulas really say or can express.

[11] Some of these labels are "traditional." We named (4) as "negative paradox" because in the standard formalization of recursive functions in combinatory logic KI is the Church numeral $\overline{0}$, while in the modeling of truth functions K stands for truth.

These formulas look weird *as implications*, because people normally seem to think that for an implication to be valid or for an inference to be logically correct it is inevitable that the sentences involved have some *information content* in common — though this is hardly a sufficient condition for validity.[12] If $\supset$ is eliminated using the rest of the connectives and the formulas are massaged through some theorems of classical logic, then the resulting expressions are much less "objectionable." However, afterward neither are they expressing inferences or implications — as the following rewriting shows.

(i) $\neg\mathcal{A} \vee \mathcal{A} \vee \mathcal{B}$ (iv) $\neg\mathcal{A} \vee \neg\mathcal{B} \vee \mathcal{B}$

(ii) $\neg\mathcal{A} \vee \mathcal{B} \vee \neg\mathcal{B}$ (v) $\neg\mathcal{A} \vee \mathcal{A} \vee \mathcal{B}$

(iii) $\neg\mathcal{A} \vee \neg\mathcal{B} \vee \mathcal{A}$ (vi) $\neg\mathcal{A} \vee \mathcal{B} \vee \neg\mathcal{B} \vee \mathcal{A}$

Essentially, either all or none of (i)–(vi) are to be accepted as valid — for instance, depending on if somebody is or is not an intuitionist (or constructivist) by his convictions.

We give one more example to illustrate the problem with classical "implication" — independently of the other connectives. Our claim is that the following sentence (when the "if. . . , then. . . "s turned into $\supset$) is true (according to classical logic) no matter if the clauses are true or false.

(7) If the world's steepest street is not far away from the University of Otago and a sculpture of Burns, then the world's steepest street is not far away from the University of Otago and a sculpture of Burns, if I purchased two books on logic in the bookstore of the University of Otago and I like milk chocolate from Cadbury's factory a few blocks from Baldwin Street.

Well, the reader might not know of my preferences in chocolate, or if I like chocolate at all; and she or he might have no idea if I ever bought any books in the bookstore at the University of Otago — though maybe a sympathetic reader supposes that I bought some logic books somewhere at some earlier time. The sentence is necessarily true (according to classical logic, of course), and not because the world's steepest street is correctly (though somewhat vaguely) located in the sentence. Furthermore, each clause alludes to locations — the University of Otago and its bookstore, Cadbury's chocolate factory, Burns's sculpture that is prominently positioned on the Octagon, and the steepest street — that are close to each other on the map. Still it is odd what my tastes and shopping habits have to do with the location of the steepest street of the world, and it is unclear why anybody would state the above sentence if only not for the purpose of *deceit*. Assume for a minute, even if it is unlikely at the moment, that development in Nepal leads to the opening of the now really steepest street in the world, which is appropriately named by the grateful Sherpas after Sir Edmund. (7) would remain true (in the sense of classical logic), and by claiming (7) one could, perhaps, give the *impression* that the steepest street is not at all in Nepal.

---

[12] The classical $\supset$ has been found problematic also from the point of view of other nonclassical logics — for example, see Barwise [1993].

The above examples, hopefully, show convincingly that classical logic's $\supset$ *does not formalize implication* — it is way too permissive. Thus, there is ample motivation to try to find a better logic that is completely rigorous and more adequate.

2 FIRST-DEGREE ENTAILMENT

An implication is relevant if the antecedent and the consequent are appropriately related. The case when there is no implication in the scope of another implication in a formula is relatively straightforward.

Let the set of formulas (wffs) be generated by a denumerable set of propositional variables ($\mathbb{P} = \{\, p_0, p_1, \dots \}$) using negation ($\sim$), conjunction ($\wedge$) and disjunction ($\vee$) in the usual way.[13] A *first-degree formula* (an **fd** wff) is two wffs $\mathcal{A}$ and $\mathcal{B}$ joined together with a binary connective $\rightarrow$, which stands for relevant implication or entailment. The idea then is to exclude the possibility that formulas like (1) and (2) are theorems or valid (of course, with $\supset$ replaced by $\rightarrow$ and $\neg$ by $\sim$), but to retain as much as possible from the usual properties of $\sim$, $\wedge$ and $\vee$. Indeed, since $\rightarrow$ can only be the main connective of an **fd** wff, the validity must depend on the other connectives. Some relevance logics — such as **B**, **T**, **E**, **R** (and the semi-relevant **RM**) — are conservative extensions of one and the same first-degree fragment: **fde**.

First-degree entailments have been characterized via normal forms, 'coupled trees' and four-valued semantics — see Anderson and Belnap [1975, §§15–16], Dunn [1976a], Dunn [1980] and Dunn [1986]. We introduce an *analytic tableaux* formulation for **fde**.[14] Analytic tableaux (also called '*semantic tableaux*') are an elegant and flexible formalism; for classical logic analytic tableaux were introduced and investigated in Smullyan [1968].

Signed formulas are usually defined as wffs prefixed with T or F; for **fde** the signs are T_i and F_i where $i \in \{\, 0, 1 \,\}$. To decide if $\mathcal{A} \rightarrow \mathcal{B}$ is a *theorem* of **fde** an **fde** *analytic tableau* is started by putting at the top $T_0\mathcal{A}, F_1\mathcal{B}$. Then the *tableau* is a tree with finite sets (singletons or doubletons) of signed formula occurrences as nodes built (top-down) according to the rules

---

[13] From now on we assume that notions such as well-formed formulas (wffs), axioms, rules, etc. are understood from the context (i.e., without the definitions repeated again and again). (For the definitions of these sorts of notions (for classical logic) — see e.g., Enderton [2001].) As far as general conventions go, $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$ range over formulas throughout and outside parentheses may be dropped.

[14] Although, perhaps, **fde** has not yet been formulated exactly as here, the **fde** tableaux are closely related to both the normal form and the coupled tree presentations.

$$\frac{T_i(\mathcal{A}\wedge\mathcal{B})}{T_i\mathcal{A},\ T_i\mathcal{B}}\ {\scriptstyle T_i\wedge}\qquad\quad \frac{T_i(\mathcal{A}\vee\mathcal{B})}{T_i\mathcal{A}\mid T_i\mathcal{B}}\ {\scriptstyle T_i\vee}\qquad\quad \frac{T_i{\sim}\mathcal{A}}{F_i\mathcal{A}}\ {\scriptstyle T_i{\sim}}$$

$$\frac{F_i(\mathcal{A}\wedge\mathcal{B})}{F_i\mathcal{A}\mid F_i\mathcal{B}}\ {\scriptstyle F_i\wedge}\qquad\quad \frac{F_i(\mathcal{A}\vee\mathcal{B})}{F_i\mathcal{A},\ F_i\mathcal{B}}\ {\scriptstyle F_i\vee}\qquad\quad \frac{F_i{\sim}\mathcal{A}}{T_i\mathcal{A}}\ {\scriptstyle F_i{\sim}},$$

where $i\in\{0,1\}$, branching is indicated by $\mid$ and (as usual) we omit $\{$ and $\}$ within the tree. We define a pair of signed formulas $T_i\mathcal{C}$ and $F_j\mathcal{C}$ to be *conjugates* whenever $i\neq j$. (Classically, $T\mathcal{C}$ and $F\mathcal{C}$ are always conjugates — there are no indices on signs or other side conditions.)

A branch of a tableau is closed when it contains conjugates, and *a tableau is closed* when all its branches are closed. Clearly, the rules shorten the formulas that are signed; therefore, a tableau can be *completed* in finitely many steps, meaning that there are no formulas in it to which a rule could be but has not yet been applied. A first-degree entailment $\mathcal{A}\to\mathcal{B}$ has a *tableaux proof* (or is *valid*) when the tableau started with $T_0\mathcal{A},F_1\mathcal{B}$ is closed. We give three simple tableaux as a quick illustration.[15]

$$\begin{array}{l}
T_0(\mathcal{A}\wedge(\mathcal{C}\vee{\sim}\mathcal{B})),\\
F_1((\mathcal{A}\wedge\mathcal{C})\vee{\sim}\mathcal{B})\\
T_0\mathcal{A},\ T_0(\mathcal{C}\vee{\sim}\mathcal{B})\\
F_1(\mathcal{A}\wedge\mathcal{C}),\ F_1{\sim}\mathcal{B}\\
\begin{array}{c|l}
\begin{array}{c} T_0\mathcal{C}\\ \begin{array}{l|l} F_1\mathcal{A} & F_1\mathcal{C}\\ * & * \end{array}\end{array} &
\begin{array}{l} T_0{\sim}\mathcal{B}\\ * \\ \end{array}
\end{array}
\end{array}
\qquad
\begin{array}{l}
T_0(p\vee(p\wedge{\sim}p\wedge q)),\\
\quad F_1p\\
\begin{array}{l|l}
T_0p & T_0(p\wedge{\sim}p\wedge q)\\
* & T_0p,\ T_0({\sim}p\wedge q)\\
 & *
\end{array}
\end{array}
\qquad
\begin{array}{l}
T_0(\mathcal{A}\wedge({\sim}\mathcal{A}\vee\mathcal{B})),\\
F_1{\sim}{\sim}\mathcal{B}\\
T_1{\sim}\mathcal{B}\\
F_1\mathcal{B}\\
T_0\mathcal{A},\ T_0({\sim}\mathcal{A}\vee\mathcal{B})\\
\begin{array}{l|l}
T_0{\sim}\mathcal{A} & T_0\mathcal{B}\\
F_0\mathcal{A} & *
\end{array}
\end{array}$$

The first two tableaux are closed; they prove that $(\mathcal{A}\wedge(\mathcal{C}\vee{\sim}\mathcal{B}))\to((\mathcal{A}\wedge\mathcal{C})\vee{\sim}\mathcal{B})$ and $(p\vee(p\wedge{\sim}p\wedge q))\to p$ are valid first-degree entailments. The third tableau is complete but open, that is, $(\mathcal{A}\wedge({\sim}\mathcal{A}\vee\mathcal{B}))\to{\sim}{\sim}\mathcal{B}$ is not valid. (This formula, with a little variation on negation, is an example of the so-called "disjunctive syllogism" in theorem form.)

The tableaux formulation provides a particularly easy decision procedure to check the validity of first-degree implications. By 'easy', of course, we do not mean the space-time *complexity*. The validity problem for the classical propositional calculus is co-NP-complete (definitely not feasible), so is it for **fde**. Instead, we mean only that the description of the procedure is brief and transparent. Actually, given a formula of classical logic with $\supset$ as its main connective one can similarly check if the formula in question corresponds to a relevant entailment by rewriting the two subformulas (on the left and on the right of $\supset$, which is, of course, also changed to $\rightarrow$) into $\sim, \wedge, \vee$ formulas. (1) and (2) (after such a transcription) are *not valid* first-degree formulas.

---

[15] Of course, if a tableau can be closed then there is no need to complete it before closing (i.e., compound conjugates necessarily yield propositional variable conjugates). We emphasize this in the examples by using propositional variables only in one of them, and closing branches, that is indicated by a $*$, as soon as possible. In order to make this chapter not too long, we rarely include proofs, let alone full proofs; sometimes we give a more or less detailed sketch, and other times (as now) only a hint (without emphasizing that details have been omitted). I hope that the reader will be able to reconstruct the missing steps or will consult the references (and otherwise, will trust me that I proved my theorems).

fde shows that it is possible to formalize implication between $\sim, \wedge, \vee$-wffs. Moreover, **fde** is robust (it has various proof systems), and stable (it has a variety of semantics). Algebraically, first-degree implications give rise to De Morgan lattices (pseudo-complemented distributive lattices). Moreover, **4** (the four element lattice in which $\sim$ has two fixed points) is characteristic for **fde** (see Dunn [1986, 182–193]). Finally, we mention that **fde** has the *perfect interpolation* property (see Anderson and Belnap [1975, 161]).

THEOREM 1. (PERFECT INTERPOLATION) *If $\mathcal{A} \rightarrow \mathcal{B}$ is an* **fde** *theorem, then there is a formula $\mathcal{C}$ such that $\mathcal{C}$ contains only propositional variables that are common to $\mathcal{A}$ and $\mathcal{B}$, and both $\mathcal{A} \rightarrow \mathcal{C}$ and $\mathcal{C} \rightarrow \mathcal{B}$ are* **fde** *theorems.*

As it is well-known Craig's interpolation theorem for classical first-order calculi (without '=') had to include additional clauses to take care of the two cases when $\neg\mathcal{A}$ and $\mathcal{B}$ are theorems. A way to avoid this problem within classical propositional logic is to introduce two constants, $\boldsymbol{T}$ for "truth" and $\boldsymbol{F}$ for "falsity," and postulate that any wff's vocabulary contains these constants (even if the formula itself does not) — see e.g., Smullyan [1968]. The perfect interpolation theorem states unconditionally that if $\mathcal{A}$ implies $\mathcal{B}$ in **fde** then the two formulas *share a propositional variable*. Of course, sharing a propositional variable, that is, sharing some *content* which can be formally described is not a sufficient condition for $\mathcal{A} \rightarrow \mathcal{B}$ to be a valid **fd** wff, however, it is a necessary one. (The first-degree formula $(\mathcal{A} \wedge (\sim\mathcal{A} \vee \mathcal{B})) \rightarrow \sim\sim\mathcal{B}$ above illustrates this point.)

fde has nothing to say about embedded implications (for instance, the analogues of (3)–(5)), thus, we proceed now to higher-degree logics.

3 **T** — TICKET ENTAILMENT IN FOCUS

For a moment let us return to informal motivations, and let us turn back to mathematical reasoning. One of the famous (and as of September 2005) open problems is the Riemann Hypothesis. Drawing an analogy with what happened to Fermat's Last Theorem, a notably easy to state and understand conjecture, it seems perfectly safe to predict that should Riemann's Hypothesis be proven (or, even if it seems unlikely, disproven), the proof will be *complicated.* Probably, there will be lemmas, subtheorems, theorems, invocations of various established facts, etc. A proof — even a not very complex proof — is a certain kind of *reasoning*, and it is rather exceptional when the premises can be arbitrarily permuted.[16] It is not customary to insert premises into mathematical proofs unless they are used in the proof. While (AC) had a not impeccable reputation in ZFC many authors took care to make explicit when a lemma or theorem was proved relying on (AC) (or any of its equivalents). Somewhat similarly Kleene [1952] flags with ° theorems of classical logic that are not intuitionistically valid. A relatively new research area called *reverse mathematics* (cf. Simpson [1991]) aims at precisely delineating the fragment of second-order Peano arithmetic Z_2 that is not only sufficient but also necessary for the proof of various (so-called) "not set-theoretical" theorems of mathematics. The observation to be made from these examples is that it is quite reasonable to limit premises to actually *used* statements. Furthermore, sometimes it makes sense to combine some of the premises together, but perhaps not certain others, then add some further premises, and in general, to proceed from the premises to the conclusion in *orderly* steps. The logic of ticket entailment imposes certain restrictions on changing the order of premises without fully disallowing permutation.

---

[16] We slightly (but intentionally) oversimplify the situation. Sometimes it is useful for purposes of exposition (even in a rigorous mathematical proof) to state and prove the "main theorem" first, and then prove the lemmas that were assumed to be true. The assumptions of a lemma

We briefly rephrase and illustrate some of the original informal motivations for **T**.[17] There is a distinction running across various disciplines between *laws* and *facts*. Disregarding for a moment epistemological reservations about how laws are discovered and justified in sciences, it is easy to recall examples of lawlike sentences from physics, for instance. Objects dropped from a certain height above sea level will accelerate toward the ground depending on the constant G (when some other factors, e.g. air resistance, are ignored). As Galilei's famous experiment proved, the dropped objects' mass is irrelevant. Statements of facts might be about particular objects dropped at particular times, for example, it might be a statement about a person performing a bungy jump in Rotorua.[18] There is usually an accompanying claim to the *primacy* of the laws over the facts; hence, the problem with facts that do not abide by the laws.

From both pictures set out above we intend to extract the idea that in reasoning it might be sensible to impose a certain, perhaps, somewhat loose *ordering* on the statements involved based on their role in the reasoning.[19]

We introduce **T**, the logic of ticket entailment, step by step, and in the process we hope to demonstrate that it is a *technically versatile logic*, which exhibits a whole range of interesting features (some of) which hold of various other relevance logics.

---

normally still would be listed together with that lemma, and not with some others; and probably, the hypotheses of one lemma would not be exchangeable with those of another theorem, etc.

[17] The name for this logic was derived from certain philosophical considerations of Ryle. See Anderson and Belnap [1975, 42], as well as Mares [2004, 43].

[18] A similar distinction between sorts of statements is drawn in many other sciences. Some might say that even in the humanities there are laws and facts, though perhaps, the laws are not quantitative.

[19] Some relevance logics mentioned in §4, for instance, **R** and **L** allow the permutation of the premises just as classical logic does. **E** restricts the permutation of premises depending on if they are implicational formulas or not, and **B** disallows all permutations.

3.1 The implicational fragment: $\mathbf{T}_{\rightarrow}$

The *pure implicational fragment* of **T** looks deceptively simple — it can be axiomatized by four axioms, (A1)–(A4), and a rule, (R1).

(A1) $\mathcal{A} \rightarrow \mathcal{A}$

(A2) $(\mathcal{A} \rightarrow \mathcal{B}) \rightarrow ((\mathcal{C} \rightarrow \mathcal{A}) \rightarrow (\mathcal{C} \rightarrow \mathcal{B}))$

(A3) $(\mathcal{A} \rightarrow \mathcal{B}) \rightarrow ((\mathcal{B} \rightarrow \mathcal{C}) \rightarrow (\mathcal{A} \rightarrow \mathcal{C}))$

(A4) $(\mathcal{A} \rightarrow (\mathcal{A} \rightarrow \mathcal{B})) \rightarrow (\mathcal{A} \rightarrow \mathcal{B})$

(R1) $\mathcal{A} \rightarrow \mathcal{B}, \mathcal{A} \Rightarrow \mathcal{B}$

The notions of a *proof* and of a *theorem* are defined in the usual way, that is, a proof is a sequence of wffs in which each formula is an axiom or is obtained from two earlier formulas by (R1); a theorem is any, but in particular, the last wff in a proof. We call these proofs *ordinary* to distinguish them from more refined notions of proofs some of which we introduce below.

Notice that this definition is very broad. Although the formulation of *modus ponens* (R1) might seem to suggest that in a detachment the implicational formula comes first, nothing actually guarantees that. Before we introduce a new notion of a proof, which also suffices to obtain all the theorems of $\mathbf{T}_{\rightarrow}$, we take a closer look at the axioms. First of all, we note that the four axioms are *principal type schemes* (in the Curry-style typing discipline) of some well-known combinators. (A1) is I's, (A2) is B's, (A3) is B′'s and (A4) is W's principal *simple type.*[20]

The laws vs facts distinction is engraved into the axioms. It does not seem sappy to assume that $\mathcal{A} \rightarrow \mathcal{A}$ should be a law; at least the following reasoning looks safe: if $\mathcal{A}$ is a fact, then we certainly may (re)state that $\mathcal{A}$ by appealing to (A1) and (R1), even though the conclusion is not new, of course. (Notwithstanding, compare this to theorem 16 (p. 751) concerning the role of the axiom (A1).) (A2) can be interpreted to say that if $\mathcal{A} \rightarrow \mathcal{B}$ and $\mathcal{C} \rightarrow \mathcal{A}$ are laws, then the fact $\mathcal{C}$ should suffice to conclude $\mathcal{B}$. (A3) is a lot alike (A2); if $\mathcal{A} \rightarrow \mathcal{B}$ and $\mathcal{B} \rightarrow \mathcal{C}$ are laws then given the fact $\mathcal{A}$ one can conclude $\mathcal{C}$. (A2) and (A3) could be summarized by saying that *implication is transitive.* Lastly, (A4) says that if a fact has been used twice to obtain a conclusion, then the repetition can be disregarded in getting to the same conclusion $\mathcal{B}$.

As a first attempt to tighten up the rather loose notion of ordinary proofs, we could require that the *major premise* of *modus ponens* (i.e., $\mathcal{A} \rightarrow \mathcal{B}$) be introduced into the proof before the *minor one* (i.e., $\mathcal{A}$). However, the resulting notion (that was introduced in Bimbó [2005a] for $\mathbf{T}_{\rightarrow}$ under the name "neat proof") is still more permissive than what we want now. The problem is that it is possible to move around formulas within a proof — some wffs can be moved toward the beginning and some wffs can be moved toward the end of the proof. Therefore, we introduce further restrictions, so that proofs contained more information about the theorem proven that ordinary proofs do. The intuitive counterpart of *traversing proofs* is a sharpening of the idea that a major premise, which is an "inference ticket" licensing an inference step, should not only precede its minor pair, and further, the minor pair of the minor pair etc., but they should come as close as possible to each other within the proof. This vague statement about how the wffs are to follow each other in a proof is made precise in the next definition.

---

[20] The axioms for these and some other combinators mentioned earlier are $\mathsf{I}x \rhd x$, $\mathsf{K}xy \rhd x$, $\mathsf{B}xyz \rhd x(yz)$, $\mathsf{B}'xyz \rhd y(xz)$, $\mathsf{W}xy \rhd xyy$.

TRAVERSING PROOFS. Let $\mathfrak{P} = \langle \mathcal{A}_1, \ldots, \mathcal{A}_z \rangle$ be a sequence of wffs, ($z \in \mathbb{Z}^+$). $\mathfrak{P}$ is a *traversing proof* if

(i) $z = 1$ and $\mathcal{A}_1$ is an instance of one of the axioms, or

(ii) $z \geq 3$ and $\mathfrak{P}$ has an initial segment $\mathfrak{P}_1 = \langle \mathcal{A}_1, \ldots, \mathcal{A}_j \rangle$ and an interior segment $\mathfrak{P}_2 = \langle \mathcal{A}_{j+1}, \ldots, \mathcal{A}_{z-1} \rangle$ (where $1 \leq j \leq z-1$) such that $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are traversing proofs, and $\mathcal{A}_j$ is $\mathcal{A}_{z-1} \to \mathcal{A}_z$.

A *theorem* is the last wff of a traversing proof.

The rationale behind (i) is that a proof which consists of two wffs — given that (R1) is the sole rule — cannot be properly understood as a proof. A timehonored observation is that the two premises of *modus ponens* cannot be identical. Therefore, in a purported proof of length two the first wff contributes nothing to the second's being a theorem. (The second wff must be an instance of an axiom, otherwise the two formulas do not constitute an (ordinary) proof.) The idea leading to (ii) is that the applications of (R1) — with the major premise coming first — might be thought of as imposing an *ordered tree* structure on a linear sequence of formulas (as long as no wffs are used repeatedly). Then (ii) reflects traversing the leaves of the proof tree from top left counter clockwise to top right. The root corresponds to the last formula in the proof.

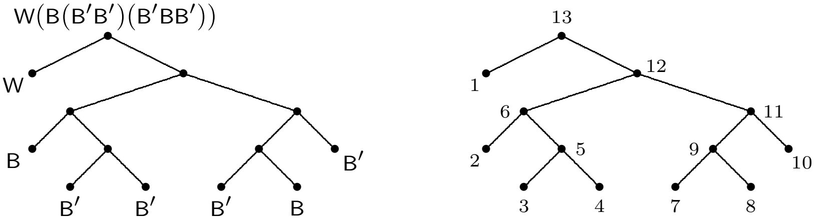


Figure 1. A combinatory term and a traversing proof

Figure 1 illustrates the similarities as well as the differences between traversing proofs and combinatory terms. The tree on the left shows the *unique formation tree* of the combinatory term $\mathsf{W}(\mathsf{B}(\mathsf{B}'\mathsf{B}')(\mathsf{B}'\mathsf{B}\mathsf{B}'))$. The tree on the right shows the indices of formulas which appear in a traversing proof of the formula $(\mathcal{A} \to \mathcal{B}) \to ((\mathcal{B} \to \mathcal{A}) \to (\mathcal{A} \to \mathcal{B}))$.[21] The example also shows that a wff may be an instance of the principal type of more that one combinator. The above wff is an instance of the principal type of the combinator K in addition to being an instance of the principal type of the combinators $\mathsf{W}(\mathsf{B}(\mathsf{B}'\mathsf{B}')(\mathsf{B}'\mathsf{B}\mathsf{B}'))$, $\mathsf{B}'(\mathsf{B}'\mathsf{B}')\mathsf{W}(\mathsf{B}'\mathsf{B}(\mathsf{B}'\mathsf{B}))$, $\mathsf{W}(\mathsf{B}(\mathsf{B}(\mathsf{B}'\mathsf{B})\mathsf{B}')\mathsf{B}')$ and so on. Of course, $(\mathcal{A} \to \mathcal{B}) \to ((\mathcal{B} \to \mathcal{A}) \to (\mathcal{A} \to \mathcal{B}))$ is *not the principal* type of K, which is not a theorem of $\mathbf{T}_\to$.

---

[21] Finding the appropriate types for the combinators, that is, instances of the axioms (A1)–(A4), is an easy exercise — given the combinatory term.

Obviously, any traversing proof is an ordinary proof; less obviously, the existence of an ordinary proof implies the existence of a traversing proof for a wff.

THEOREM 2. (SETS OF THEOREMS) *A formula $\mathcal{A}$ has an* ordinary proof *if and only if $\mathcal{A}$ has a* traversing proof. *A provable formula has $\aleph_0$ ordinary proofs.*

Tidying up an ordinary proof into a traversing proof may require deleting, inserting and rearranging formulas. Given a traversing proof it is sufficient to choose one axiom and generate its instances to be inserted one by one as the first wff into the proof to get $\aleph_0$ ordinary proofs.

The set of traversing proofs is a *proper* subset of the set of ordinary proofs, and in this sense traversing proofs are more restrictive, though they do not change the set of theorems. The correspondence between combinatory terms and traversing proofs provides a better grasp of $\mathbf{T}_\to$ theorems (cf. §3.4). Notwithstanding, traversing proofs can replace the notion of ordinary proofs in other logics too, such as $\mathbf{E}_\to$, $\mathbf{L}_\to$ and $\mathbf{R}_\to$ (cf. §4). Thus, one can also look at the "inference ticket idea" as a more rigorous way to *analyze proofs.*

Proofs are often defined as derivations from an empty set of premises, thus, a natural question is if it is possible to define deductions similarly to proofs. First of all, we observe that if deductions in $\mathbf{T}_\to$ were defined like ordinary proofs, then the analogue of the deduction theorem would turn out to be false — trivially. However, there are various ways to define deductions (with an appropriate deduction theorem) from traversing proofs. A not too complicated definition is as follows.

T-DEDUCTIONS. Let $\langle \mathcal{A}_1, \ldots, \mathcal{A}_j, \mathcal{B}_1, \ldots, \mathcal{B}_z \rangle$ be a string of wffs ($j, z \in \mathbb{Z}^+$) such that (i) $\langle \mathcal{A}_1, \ldots, \mathcal{A}_j \rangle$ is a traversing proof of $\mathcal{A}_j$, and (ii) for all $\langle \mathcal{B}_{2i+1}, \mathcal{B}_{2i+2} \rangle$ (where $i \in \mathbb{Z}$, $0 \leq 2i \leq z-2$) $\mathcal{B}_{2i+2}$ is obtained by detaching $\mathcal{B}_{2i+1}$ from the immediately preceding wff in the string. Then $\mathcal{A}_1, \ldots, \mathcal{A}_j, \mathcal{B}_1, \ldots, \mathcal{B}_{z-1} \vdash_{\mathbf{T}_\to} \mathcal{B}_z$, that is, there is a **T**-*deduction* of $\mathcal{B}_z$ from $\mathcal{A}_1, \ldots, \mathcal{A}_j, \mathcal{B}_1, \ldots, \mathcal{B}_{z-1}$.

THEOREM 3. (**T**-DEDUCTION THEOREM) *If $\mathcal{A}_1, \ldots, \mathcal{A}_j, \mathcal{B}_1, \ldots, \mathcal{B}_{z-1} \vdash_{\mathbf{T}_\to} \mathcal{B}_z$ then there is a* **T**-deduction *of $\mathcal{B}_{z-1} \to \mathcal{B}_z$ from $\mathcal{A}_1, \ldots, \mathcal{A}_j, \mathcal{B}_1, \ldots, \mathcal{B}_{z-2}$ or $\mathcal{B}_{z-1} \to \mathcal{B}_z$ is a* theorem.

The proof of this theorem is practically obvious. The results of §3.4 reveal that it is possible to define a more sophisticated notion of deduction for $\mathbf{T}_\to$ without loosing either the deduction theorem or the relevance of deductions. A deduction theorem for $\mathbf{R}_\to$ was proven by Church and Moh. See Dunn [1986, §1.4].

$\mathbf{T}_\to$ demonstrates that good properties of implication — such as transitivity, for instance — can be formalized *without a recourse to classical logic.* Although we have not proven (yet) that $\mathbf{T}_\to$ is a proper sublogic of (that is, not identical to) the "strong implicational calculus" with Peirce's law, it could be proven in various ways.[22] (1) For instance, there is a well-known correspondence between the provability of Peirce's law and a sequent calculus having multiple right hand sides (without restricted implication introduction on the right). Thus, if a suitable Gentzenization can be found for $\mathbf{T}_{\rightarrow}$ with single right hand side, then this can demonstrate that $\mathbf{T}_{\rightarrow}$ cannot prove Peirce's law (cf. §3.2, p. 741). (2) Another way to argue the same point is to appeal to the correspondence between combinatory terms and theorems. Every theorem of $\mathbf{T}_{\rightarrow}$ is a principal type scheme of some combinator over the basis $\{\mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{W}\}$. For example, K and C are not definable from this basis.[23] Therefore, $\mathbf{T}_{\rightarrow}$ does not prove $(\mathcal{A} \rightarrow (\mathcal{B} \rightarrow \mathcal{C})) \rightarrow (\mathcal{B} \rightarrow (\mathcal{A} \rightarrow \mathcal{C}))$ either, although it is easily seen to be a theorem of classical logic. There is a similar correspondence between simple types — axiom schemes — and combinatory bases, in particular, between $\mathbf{E}_{\rightarrow}$ and $\{\mathsf{I}, \mathsf{B}', \mathsf{C}(\mathsf{B}(\mathsf{CC})(\mathsf{B}'\mathsf{I})), \mathsf{W}\}$ and $\mathbf{R}_{\rightarrow}$ and $\{\mathsf{I}, \mathsf{B}, \mathsf{C}, \mathsf{W}\}$ — see §4. (3) Yet another way to prove that $\mathbf{T}_{\rightarrow}$ does not collapse into classical logic is to define a semantics that is sound for $\mathbf{T}_{\rightarrow}$, but not for classical logic (cf. §3.3).

In sum, $\mathbf{T}_{\rightarrow}$ shows that a reasonable and rigorously formalizable notion of implication need not to collapse into classical logic. Moreover, this logic has straightforward connections to other areas of mathematical and computational logic. One might ask if we have lost something (beyond the connectives $\sim$, $\vee$ and $\wedge$, which were not included in $\mathbf{T}_{\rightarrow}$). Classical logic and **fde** are decidable (cf. p. 730), but no similar procedure seems to exist for $\mathbf{T}_{\rightarrow}$ — perhaps, decidability is lost. (Unless otherwise specified by "decidability" we mean the decidability of *theoremhood*, or equivalently the decidability of validity, if applicable.) The problematic axiom is (A4), so to speak. This axiom together with (A2) and (A3) is the one that causes difficulty when one tries to define a decision procedure in a more or less direct way. This observation is underscored by the fact that the logic $\mathbf{T}\text{-}\mathbf{W}_{\rightarrow}$ is decidable (cf. Anderson *et al.* [1992, §67] as well as theorem 9 below). Adding more permutation than what is contained in (A3) also helps to establish decidability: $\mathbf{E}_{\rightarrow}$ and $\mathbf{R}_{\rightarrow}$ (which both include contraction) are decidable just as $\mathbf{E}_{\underset{\rightarrow}{\sim}}$ and $\mathbf{R}_{\underset{\rightarrow}{\sim}}$ are — cf. Kripke [1959], Anderson and Belnap [1975, §13] and Dunn [1986, §§3.6–7].

Axiom (A4) may be replaced by axiom (A5), which is closely related to (A6), the "self-distribution of implication."

(A5) $(\mathcal{A} \rightarrow \mathcal{B}) \rightarrow ((\mathcal{A} \rightarrow (\mathcal{B} \rightarrow \mathcal{C})) \rightarrow (\mathcal{A} \rightarrow \mathcal{C}))$

(A6) $(\mathcal{A} \rightarrow (\mathcal{B} \rightarrow \mathcal{C})) \rightarrow ((\mathcal{A} \rightarrow \mathcal{B}) \rightarrow (\mathcal{A} \rightarrow \mathcal{C}))$

That is, the implicational fragment of ticket entailment may be defined equivalently by (R1), (A1)–(A3) and (A5). ((A6) frequently appears as an axiom of classical logic, but it is a theorem of $\mathbf{T}_{\rightarrow}$ too, and so it can be added as an axiom if desired.) (A5) and (A6) are the principal type schemes of S' and S, respectively. The axioms of these combinators are $\mathsf{S}xyz \rhd xz(yz)$ and $\mathsf{S}'xyz \rhd yz(xz)$. ($\mathsf{W}$ is definable as $\mathsf{S}'\mathsf{I}$, on the other hand, $\mathsf{B}(\mathsf{B}(\mathsf{B}'\mathsf{B}')(\mathsf{BW}))\mathsf{B}'$ is S and $\mathsf{B}(\mathsf{BW})(\mathsf{BBB}')$ is S'.)

---

[22] $\mathbf{T}_{\rightarrow}$ is, obviously, a "sublogic" of classical logic, because (A1)–(A4) (with $\supset$ instead of $\rightarrow$, of course) are theorems of and *modus ponens* is a rule of classical logic.

[23] The axiom of C is $\mathsf{C}xyz \rhd xzy$; for axioms of some other combinators (including K's) see footnote 20 (p. 733).

3.2 Intensional fragments: $\mathbf{LT}^{t}_{\to}$*,* $\mathbf{LT}^{\circ t}_{\to}$*,* $\mathbf{T}^{\circ t}_{\to}$*,* $\mathbf{T}^{t}_{\to}$ *and* $\mathbf{LT}_{\to}$

$\mathbf{T}_{\to}$ can be defined as a "Gentzen-style" calculus, however, we have to take some preparatory steps. In Gentzen's sequent calculi **LK** and **LJ**, which formalize classical and intuitionistic predicate logics, there are connectives beyond $\supset$, in particular, there is conjunction and disjunction. Furthermore, the Lindenbaum algebra of classical logic (and also that of intuitionistic logic) is based on an implicative lattice, (i.e., $\supset$ is the *residual* of $\wedge$). Conjunction is commutative and a conjunction implies its conjuncts. If $\to$ were taken to be the residual of $\wedge$ then the positive and negative paradoxes, (3) and (4), would reappear immediately. Therefore, a structural connective in a sequent calculus must have a different relationship to $\wedge$ than , has.

The first sequent calculus $L\mathbf{R}^{+}$ that contained a second structural connective was introduced in 1969 by Dunn (cf. Dunn [1973]). In §3.5 we will use two structural connectives in $\mathbf{LT}^{\circ t}_{+}$, however, at the moment we only introduce ;. Informally, ; corresponds to putting premises together. Unlike , , which is a polyadic structural connective, ; is a binary structural connective. Moreover, ; is not assumed to be associative or commutative. To be able to define the *structural rules* appropriately we need a notion between a formula and an antecedent of a sequent. "Bunches" of formulas will be called *structures.*

STRUCTURES. The *set of structures*, denoted by `str`, is defined as follows.

(i) If $\mathcal{A}$ is a wff then $\mathcal{A} \in$ `str`;

(ii) if $\mathfrak{A}, \mathfrak{B} \in$ `str` then $(\mathfrak{A}; \mathfrak{B}) \in$ `str`.

Note that the base set of the induction is the set of wffs and there is no empty structure in `str`. ($\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \ldots$ range over (nonempty) structures everywhere.) Thus, no empty left (or right) hand sides are allowed in our sequent calculi (unlike in **LK** or **LJ**).

A structure can be *uniquely decomposed* into *substructures.* As a consequence, occurrences of structures in structures can be identified unambiguously. We do not need here detailed conventions for the identification of arbitrary substructures or formulas occurring in structures; we need only that such an identification is possible.

$\mathfrak{B}[\mathfrak{A}]$ is the structure $\mathfrak{B}$ in which one particular occurrence of $\mathfrak{A}$ is singled out.[24] Additionally, we assume that this bracket notation allows the identification of occurrences across limited contexts such as within a rule. Then, if $\mathfrak{C}$ is a structure and $\mathfrak{B}[\mathfrak{C}/\mathfrak{A}]$ is the result of replacing the chosen occurrence of $\mathfrak{A}$ in $\mathfrak{B}$ by $\mathfrak{C}$ then the result of this replacement is also denoted by $\mathfrak{B}[\mathfrak{C}]$ (when there is no risk of confusion). The notation can be applied repeatedly to a structure $\mathfrak{B}$; if the structures in the brackets are all the same we simply replace the square brackets with angle ones. That is, $\mathfrak{A}\langle\mathfrak{B}\rangle$ is the structure $\mathfrak{A}$ in which at least one, but possibly several occurrences of $\mathfrak{B}$ have been fixed.

---

[24] In λ-calculus $t[\ ]$ is sometimes used to indicate a *context* (or a "hole") in the term t — which is not very far in spirit from what is meant by $\mathfrak{B}[\ ]$.

For technical reasons — essentially, to make some proofs easier — we add a constant (a zeroary connective) to the logic, that is denoted by $\boldsymbol{t}$. Informally, $\boldsymbol{t}$ is *the truth*, that is, the greatest lower bound of all theorems. $\boldsymbol{t}$ has been added to both consecution and axiomatic calculi before (see Anderson *et al.* [1992, §28]), and $\boldsymbol{t}$ is of chief importance in Gentzenizations of positive fragments of **R**, **E**, **T** etc. — see Dunn [1986, §3.10].[25]

SEQUENT CALCULUS $\mathbf{LT}^{\boldsymbol{t}}_{\to}$. *Sequents* (or *consecutions*) are formed from a structure, $\vdash$ and a formula. $\mathbf{LT}^{\boldsymbol{t}}_{\to}$ is defined by the following axiom and rules.

$$\mathcal{A}\vdash\mathcal{A}\ \ \mathsf{I}\qquad \frac{\mathfrak{A}\vdash\mathcal{A}\quad \mathfrak{B}[\mathcal{B}]\vdash\mathcal{C}}{\mathfrak{B}[\mathcal{A}\to\mathcal{B};\mathfrak{A}]\vdash\mathcal{C}}\ {\to}{\vdash}\qquad \frac{\mathfrak{A};\mathcal{A}\vdash\mathcal{B}}{\mathfrak{A}\vdash\mathcal{A}\to\mathcal{B}}\ {\vdash}{\to}$$

$$\frac{\mathfrak{A}[\mathfrak{B};(\mathfrak{C};\mathfrak{D})]\vdash\mathcal{A}}{\mathfrak{A}[(\mathfrak{B};\mathfrak{C});\mathfrak{D}]\vdash\mathcal{A}}\ \mathsf{B}{\vdash}\qquad \frac{\mathfrak{A}[\mathfrak{C};(\mathfrak{B};\mathfrak{D})]\vdash\mathcal{A}}{\mathfrak{A}[(\mathfrak{B};\mathfrak{C});\mathfrak{D}]\vdash\mathcal{A}}\ \mathsf{B}'{\vdash}\qquad \frac{\mathfrak{A}[(\mathfrak{B};\mathfrak{C});\mathfrak{C}]\vdash\mathcal{A}}{\mathfrak{A}[\mathfrak{B};\mathfrak{C}]\vdash\mathcal{A}}\ \mathsf{W}{\vdash}$$

$$\frac{\mathfrak{A}[\mathfrak{B}]\vdash\mathcal{A}}{\mathfrak{A}[\boldsymbol{t};\mathfrak{B}]\vdash\mathcal{A}}\ \boldsymbol{t}{\vdash}\qquad \frac{\mathfrak{A}[\boldsymbol{t};\boldsymbol{t}]\vdash\mathcal{A}}{\mathfrak{A}[\boldsymbol{t}]\vdash\mathcal{A}}\ \mathsf{M}_{\boldsymbol{t}}{\vdash}$$

A proof is a certain kind of tree (exactly as in other sequent calculi); and a formula $\mathcal{A}$ is a *theorem* of $\mathbf{LT}^{\boldsymbol{t}}_{\to}$ when $\boldsymbol{t}\vdash\mathcal{A}$ has a proof. The labels for the structural rules (except, perhaps, ($\mathsf{M}_{\boldsymbol{t}}\vdash$)) are hopefully self-explanatory based on our previous remarks.[26] The subscript in the label of the rule ($\mathsf{M}_{\boldsymbol{t}}\vdash$) shows that for this rule to be applicable two structures have to be $\boldsymbol{t}$.

The piecemeal *single* cut and the *multiple* cut rules are as follows.

$$\frac{\mathfrak{A}\vdash\mathcal{C}\quad \mathfrak{B}[\mathcal{C}]\vdash\mathcal{A}}{\mathfrak{B}[\mathfrak{A}]\vdash\mathcal{A}}\ \text{cut}\qquad\qquad \frac{\mathfrak{A}\vdash\mathcal{C}\quad \mathfrak{B}\langle\mathcal{C}\rangle\vdash\mathcal{A}}{\mathfrak{B}\langle\mathfrak{A}\rangle\vdash\mathcal{A}}\ \text{cut}$$

The two rules bear the same label, because they are equal in power.

THEOREM 4. (CUTS ARE EQUIPOTENT) *The* single *and the* multiple cut *rule allow the* same *sequents to be proven.*

In the presence of (W$\vdash$) certain proof techniques (derived from Gentzen's proof) fail to ensure the admissibility of single cut. However, the usual adaptation of Gentzen's double induction proof of the eliminability of mix fails to demonstrate the admissibility of multiple cut in $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$. (We denote by $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$ $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$ without $(\mathsf{W}\vdash)$.)

---

[25] $\mathbf{LT}^{\boldsymbol{t}}_{\to}$ below is similar to the $\to$, $\boldsymbol{t}$-fragment of the $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$-**W** calculus of Giambrone (cf. Anderson *et al.* [1992, §67.2]). The differences, however, go beyond the obvious ones such as omitting the extensional connectives and adding a rule for contraction.

[26] M is definable as WI and (following Curry) M is occasionally denoted by W_*. The axiom for M is $\mathsf{M}x \rhd xx$. Incidentally, notice that the set of combinators that have a simple type is *not closed* under combinatory application, because there is no $\to$ wff corresponding to M.

THEOREM 5. (ADMISSIBILITY OF THE CUT) *The multiple cut rule is* admissible *in* $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$ *and* $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$.

The theorem may be proven by an application of the proof technique introduced in Bimbó [2005b].

Corollary. (ADMISSIBILITY OF SINGLE CUT RULE) *The* single cut rule *is* admissible *in* $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$ *as well as in* $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$.

In **LK** (or in **LJ**) all provable sequents are equivalent to formulas; not so in $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$ (or in $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$), because ; lacks certain properties that , has. Now we introduce a connective that enables one to turn any structures into wffs, therefore, all provable sequents into provable formulas. If the antecedent of a consecution is left associated with $\boldsymbol{t}$ as the leftmost structure, then all the wffs (save $\boldsymbol{t}$) can be pushed across the turnstile into an implicational formula. Otherwise, the binary connective $\circ$ (called *fusion* or *tensor product*) can yield a formula.

FUSION ADDED TO SEQUENT CALCULI $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$ AND $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$. The logics $\mathbf{LT}^{\circ\boldsymbol{t}}_{\rightarrow}$ and $\mathbf{LT\text{-}W}^{\circ\boldsymbol{t}}_{\rightarrow}$ are defined by adding the rules $(\circ\vdash)$, $(\vdash\circ)$ and $(\mathsf{b}_{\boldsymbol{t}}\vdash)$ to $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$ and $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$.[27]

$$\frac{\mathfrak{A}[\mathcal{A};\mathcal{B}]\vdash\mathcal{C}}{\mathfrak{A}[\mathcal{A}\circ\mathcal{B}]\vdash\mathcal{C}}\,\circ\vdash \qquad \frac{\mathfrak{A}\vdash\mathcal{A}\quad\mathfrak{B}\vdash\mathcal{B}}{\mathfrak{A};\mathfrak{B}\vdash\mathcal{A}\circ\mathcal{B}}\,\vdash\circ \qquad \frac{\mathfrak{A}[(\boldsymbol{t};\mathfrak{B});\mathfrak{C}]\vdash\mathcal{A}}{\mathfrak{A}[\boldsymbol{t};(\mathfrak{B};\mathfrak{C})]\vdash\mathcal{A}}\,\mathsf{b}_{\boldsymbol{t}}\vdash$$

The subscript in $(\mathsf{b}_{\boldsymbol{t}}\vdash)$ indicates that not all structures are arbitrary, in this case only one has to be $\boldsymbol{t}$. $(\mathsf{b}_{\boldsymbol{t}}\vdash)$ — even in the absence of $(\mathsf{M}_{\boldsymbol{t}}\vdash)$ — causes a problem for a Gentzen-style proof of ET. However, once again it is possible to use four inductions to prove the admissibility of the cut rule.

THEOREM 6. (ADMISSIBILITY OF CUT RULE) *The cut rule is* admissible *in* $\mathbf{LT}^{\circ\boldsymbol{t}}_{\rightarrow}$ *and* $\mathbf{LT\text{-}W}^{\circ\boldsymbol{t}}_{\rightarrow}$.

The admissibility of the cut rule has various pleasant consequences such as the *subformula property.* In $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$, $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$, $\mathbf{LT}^{\circ\boldsymbol{t}}_{\rightarrow}$ and $\mathbf{LT\text{-}W}^{\circ\boldsymbol{t}}_{\rightarrow}$ this means that if there is a proof of a sequent $\mathfrak{A}\vdash\mathcal{A}$, then there is a proof such that the root sequent $\mathfrak{A}\vdash\mathcal{A}$ contains all formulas (perhaps, as subformulas) that occur in a consecution in the proof. This way of putting the subformula property emphasizes the role of the above theorem as a cut *elimination theorem* (when, to begin with, the calculi are defined to include the cut rule). An inspection of the connective and structural rules verifies that any formula in the upper sequent of any rule enters the lower sequent as a formula or a subformula; therefore, the subformula property implies the truth of the next theorem.

THEOREM 7. (CONSERVATIVITY OF FUSION) $\mathbf{LT}^{\circ\boldsymbol{t}}_{\rightarrow}$ *and* $\mathbf{LT\text{-}W}^{\circ\boldsymbol{t}}_{\rightarrow}$ *are* conservative extensions *of* $\mathbf{LT}^{\boldsymbol{t}}_{\rightarrow}$ *and* $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$, *respectively.*

---

[27] b is a *dual combinator*; its axiom is $x(y(z\mathsf{b}))\rhd xyz$. b does not have a → type, although it has a ← type, viz., $((\mathcal{A}\leftarrow\mathcal{B})\leftarrow(\mathcal{C}\leftarrow\mathcal{B}))\leftarrow(\mathcal{A}\leftarrow\mathcal{C})$ in an extended type assignment system.

The inclusion of fusion makes the proof of the equivalence of sequent and axiomatic calculi straightforward. The axiomatic calculi $\mathbf{T}^{\circ\boldsymbol{t}}_{\to}$ and $\mathbf{T\text{-}W}^{\circ\boldsymbol{t}}_{\to}$ are defined by (A1)–(A4) and (A1)–(A3), together with rules (R1)–(R5). The new rules are

(R2) $\vdash \mathcal{A} \to (\mathcal{B} \to \mathcal{C}) \;\Rightarrow \vdash (\mathcal{A} \circ \mathcal{B}) \to \mathcal{C}$

(R3) $\vdash (\mathcal{A} \circ \mathcal{B}) \to \mathcal{C} \;\Rightarrow \vdash \mathcal{A} \to (\mathcal{B} \to \mathcal{C})$

(R4) $\vdash \mathcal{A} \;\Rightarrow \vdash \boldsymbol{t} \to \mathcal{A}$

(R5) $\vdash \boldsymbol{t} \to \mathcal{A} \;\Rightarrow \vdash \mathcal{A}$

If $\mathfrak{A} \vdash \mathcal{A}$ is a provable sequent then $\boldsymbol{t} \vdash \mathfrak{A}^\circ \to \mathcal{A}$ is provable too (where $\mathfrak{A}^\circ$ is $\mathcal{B}$ if $\mathfrak{A}$ is $\mathcal{B}$, and otherwise $\mathfrak{A}^\circ$ is $\mathfrak{B}^\circ \circ \mathfrak{C}^\circ$ if $\mathfrak{A}$ is $\mathfrak{B}; \mathfrak{C}$). Conversely, if $\boldsymbol{t} \vdash \mathfrak{A}^\circ \to \mathcal{A}$ is provable and $\mathcal{A}$ is not $\boldsymbol{t} \circ \mathcal{B}$ then $\mathfrak{A} \vdash \mathcal{A}$ is provable as well.

THEOREM 8. (Equivalence of consecution and axiomatic calculi) *The sequent* $\boldsymbol{t} \vdash \mathcal{A}$ *is* provable *in* $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$ *(or* $\mathbf{LT\text{-}W}^{\circ\boldsymbol{t}}_{\to}$*) if and only if* $\mathcal{A}$ *is a* theorem *of* $\mathbf{T}^{\circ\boldsymbol{t}}_{\to}$ *(or, respectively, of* $\mathbf{T\text{-}W}^{\circ\boldsymbol{t}}_{\to}$*).*

The proof, going from the axiomatic to the consecution calculi, is straightforward. We only point out that $(\mathsf{M}_{\boldsymbol{t}}\vdash)$ is useful in proving (R1) and (R5), while $(\mathsf{b}_{\boldsymbol{t}}\vdash)$ is utilized in proving (R2).

Sequent calculi sometimes can be used to establish the decidability of a logic — more or less similarly to the proofs of the decidability of classical and intuitionistic propositional logics. Using the equivalence of the consecution and axiomatic calculi, together with both $\mathbf{LT\text{-}W}^{\circ\boldsymbol{t}}_{\to}$ and $\mathbf{T\text{-}W}^{\circ\boldsymbol{t}}_{\to}$ conservatively extending their counterpart $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\to}$ and $\mathbf{T\text{-}W}^{\boldsymbol{t}}_{\to}$, we can state a completely straightforward decidability result. (The decidability of $\mathbf{T\text{-}W}_{\to}$ follows from Giambrone's theorem for $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}\mathbf{\text{-}W}$ — see Anderson *et al.* [1992, §67]. We give a sketch of a different proof.)

THEOREM 9. (Decidability of $\mathbf{T\text{-}W}_{\to}$) *The logic* $\mathbf{T\text{-}W}_{\to}$ *is* decidable.

Proof. First one proves that $\boldsymbol{t}$ is conservatively added to $\mathbf{T\text{-}W}_{\to}$ by the rules (R4) and (R5). Then, from the equivalence of the logics follows that $\vdash_{\mathbf{T\text{-}W}^{\boldsymbol{t}}_{\to}} \mathcal{A} \to \mathcal{B}$ if and only if $\boldsymbol{t} \vdash_{\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\to}} \mathcal{A} \to \mathcal{B}$. Since $\boldsymbol{t}$ does not occur in $\mathcal{B}$, all the occurrences in the derivation of $\boldsymbol{t} \vdash \mathcal{A} \to \mathcal{B}$ must have been introduced by $(\boldsymbol{t}\vdash)$. Furthermore, no $(\to\vdash)$ rule application in the proof can have $\boldsymbol{t} \vdash \mathcal{C}$ as its left premise, because the structural rules do not allow $\boldsymbol{t}$ to be permuted into its final position; let alone can $\boldsymbol{t} \vdash \mathcal{C}$ be the right premise. Notice that $\mathbf{LT\text{-}W}^{\boldsymbol{t}}_{\to}$ has not only the usual subformula property, but a stronger one: no *formula occurrences*, with the exception of occurrences of $\boldsymbol{t}$, can disappear in a proof. Since the only remaining role for $\boldsymbol{t}$ is to ensure that the left hand side of $\vdash$ does not become empty, in a backward proof search it is enough to keep one $\boldsymbol{t}$ in the sequent. This brings about a decision procedure. ■

Parity of propositional variables in $\mathbf{T\text{-}W}_{\to}$. An immediate consequence of the *strong subformula property* is that any propositional variable occurring in a theorem of $\mathbf{T\text{-}W}_{\to}$ occurs *evenly many times.*

This observation excludes the provability in $\mathbf{T\text{-}W}_{\to}$ of the formula (A4) — which is, of course, the axiom that has been dropped — but also that of (A5) and (A6) (s. p. 736), and of some further formulas from among those that are principal types of combinators with duplicative effect. A subclass of the theorems of $\mathbf{T\text{-}W}_{\to}$ has a sharper characterization. (Some details of the proof are in Anderson *et al.* [1992, §66].)

THEOREM 10. (MUTUAL PROVABILITY IN $\mathbf{T\text{-}W}_{\to}$ (MARTIN, 1978)) *If both* $\mathcal{A} \to \mathcal{B}$ *and* $\mathcal{B} \to \mathcal{A}$ *are theorems of* $\mathbf{T\text{-}W}_{\to}$ *then* $\mathcal{A}$ *and* $\mathcal{B}$ *are the* same formula.

A rather interesting question is if it is possible to define a consecution calculus for the pure implicational part of ticket entailment without extending the set of logical connectives, that is, without the addition of fusion and truth. (It is possible to so Gentzenize some other relevance logics, e.g., $\mathbf{R}_{\to}$ and $\mathbf{R}_{\overset{\sim}{\to}}$ — see Dunn [1986, §§3.6–7].) Of course, fusion can be gotten rid off at once since $\mathbf{LT}^{\circ \boldsymbol{t}}_{\to}$ is a conservative extension over $\mathbf{LT}^{\boldsymbol{t}}_{\to}$. Thus the question is simplified to if it is possible to eliminate $\boldsymbol{t}$. The answer — somewhat surprisingly — is *yes.*

First we define $\mathbf{LT}_{\to}$, that is going to include some but not all of the rules of $\mathbf{LT}^{\circ \boldsymbol{t}}_{\to}$, then we prove that it suffices.

SEQUENT CALCULUS FOR IMPLICATIONAL TICKET ENTAILMENT. $\mathbf{LT}_{\to}$ contains the axiom (I), and the rules ($\to\vdash$), ($\vdash\to$), ($\mathsf{B}\vdash$), ($\mathsf{B}'\vdash$), ($\mathsf{W}\vdash$) and ($\mathsf{M}\vdash$) where the last rule is the unrestricted variant of ($\mathsf{M}_{\boldsymbol{t}}\vdash$).

Cut is *admissible* in $\mathbf{LT}_{\to}$, moreover, this can be shown (for multiple cut) by the simple (Gentzen-style) double induction argument on the degree and rank of the cut formula. No empty left hand sides are allowed, which suggests that we should define $\mathcal{A} \to \mathcal{B}$ to be provable in $\mathbf{LT}_{\to}$ when $\mathcal{A} \vdash \mathcal{B}$ is provable. Of course, all theorems of $\mathbf{T}_{\to}$ are implicational formulas, and it is easy to prove (indeed, parts of some previously mentioned proofs suffice here) that all of (A1)–(A4) are provable in $\mathbf{LT}_{\to}$. However, a difficulty arises if one tries in a direct way simulate *modus ponens* in this sequent calculus. One would need to show that $\mathcal{A}_1 \vdash \mathcal{A}_2$ and $\mathcal{A}_1 \to \mathcal{A}_2 \vdash \mathcal{B}$ implies that $\mathcal{B}$ is $\mathcal{B}_1 \to \mathcal{B}_2$ for some wffs $\mathcal{B}_1$ and $\mathcal{B}_2$, and moreover, that $\mathcal{B}_1 \vdash \mathcal{B}_2$ is provable. Perhaps, one can find a way to prove this claim directly (possibly, using a cut or cuts). Of course, in $\mathbf{T}_{\to}$ $\vdash \mathcal{A}$ and $\vdash \mathcal{A} \to \mathcal{B}$ are easily seen to imply — due to $\mathbf{T}$'s consistency — that $\mathcal{B}$ is $\mathcal{B}_1 \to \mathcal{B}_2$. However, even if the assumption that $\mathcal{B}$ is an implication is granted, still it has to be proven that the provability of $\mathcal{A}_1 \to \mathcal{A}_2; \mathcal{B}_1 \vdash \mathcal{B}_2$ implies the provability of $\mathcal{B}_1 \vdash \mathcal{B}_2$. It seems to us that an argument establishing that $\mathcal{A}_1 \to \mathcal{A}_1$ is "superfluous" would go through a similar reasoning that we give below exploiting the connection between $\mathbf{LT}_{\to}$ and $\mathbf{LT}^{\boldsymbol{t}}_{\to}$. In any case, what needs to be proven is not the usual match between cut and detachment, therefore, we argue differently.

$\mathbf{LT}^{\boldsymbol{t}}_{\to}$ proves exactly the same $\to, \boldsymbol{t}$-theorems as $\mathbf{T}^{\boldsymbol{t}}_{\to}$ does, since $\mathbf{T}^{\boldsymbol{t}}_{\to}$ is a conservative extension of $\mathbf{T}_{\to}$ (cf. Anderson *et al.* [1992]). For any theorem $\mathcal{A}$ of $\mathbf{T}_{\to}$ (i.e., for any $\mathcal{A}$ such that $\vdash_{\mathbf{T}_{\to}} \mathcal{A}$), $\vdash_{\mathbf{T}^{\boldsymbol{t}}_{\to}} \mathcal{A}$, hence, $\boldsymbol{t} \vdash_{\mathbf{LT}^{\boldsymbol{t}}_{\to}} \mathcal{A}$. Since the cut rule is admissible (or for the present point we should stress eliminable), there is a *cut-free* *proof* of this sequent in $\mathbf{LT}^{\boldsymbol{t}}_{\to}$. $\mathcal{A}$ is by necessity of the form $\mathcal{B} \to \mathcal{C}$ where neither $\mathcal{B}$ nor $\mathcal{C}$ has $\boldsymbol{t}$ as its subformula. The proof's last connective rule is $(\vdash\to)$. Since $\boldsymbol{t}$ is not a subformula of $\mathcal{B} \to \mathcal{C}$, $\boldsymbol{t}$ must have been introduced by an application of $(\boldsymbol{t}\vdash)$. Furthermore, due to its position in the final sequent, $\boldsymbol{t}$ could not be the whole $\mathfrak{A}$ in an application of $(\to\vdash)$. $\boldsymbol{t}$ plays no special role in $(\vdash\to)$ steps, in particular, $\boldsymbol{t}$ can be neither $\mathcal{B}$ nor the all $\mathfrak{B}[\mathcal{B}]$ of the $(\to\vdash)$ rule, because $\boldsymbol{t}$ is not a subformula of the theorem we are considering. Therefore, any $(\to\vdash)$ step in $\mathbf{LT}^{\boldsymbol{t}}_{\to}$ can be repeated (or simulated) in $\mathbf{LT}_{\to}$. Any $(\mathsf{B}\vdash)$ or $(\mathsf{B}'\vdash)$ step can be either repeated without $\boldsymbol{t}$, or if $\boldsymbol{t}$ by itself is one of the structures $\mathfrak{B}, \mathfrak{C}$ or $\mathfrak{D}$ then the step may be completely omitted. $\boldsymbol{t}$ cannot form by itself the duplicated structure in the $(\mathsf{W}\vdash)$ rule because it is not a right identity. It only remains to consider when the $(\mathsf{W}\vdash)$ rule could be applied because $\boldsymbol{t}$ was the whole structure $\mathfrak{B}$. However, after applying $(\boldsymbol{t}\vdash)$ backwards, now $(\mathsf{M}\vdash)$ can be applied to yield the same sequent. Lastly, we note that $(\mathsf{M}\vdash)$ is an admissible rule in $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$, therefore, by inclusion, it does not lead to the proof of non-theorems of $\mathbf{T}_{\to}$ in $\mathbf{LT}_{\to}$ either.

It seems to us that beyond the *elegance* of $\mathbf{LT}_{\to}$, this consecution calculus might be useful in obtaining further technical results about $\mathbf{T}_{\to}$, since it provides a tight grasp on sequents without the burden of extra vocabulary.

Now we turn back to $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$ to show that this is a calculus that is easy to handle. $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$ and $\mathbf{LT}\text{-}\mathbf{W}^{\circ\boldsymbol{t}}_{\to}$ can be "algebraized" by defining an equivalence relation between formulas as follows. Two wffs $\mathcal{A}$ and $\mathcal{B}$ are *equivalent* whenever both $\boldsymbol{t} \vdash \mathcal{A} \to \mathcal{B}$ and $\boldsymbol{t} \vdash \mathcal{B} \to \mathcal{A}$ are provable. The proof that the so defined relation is really an equivalence relation and a *congruence* (when the operations are "lifted" to equivalence classes) can be easily carried out via appealing to $(\mathsf{B}\vdash)$, $(\mathsf{B}'\vdash)$, $(\mathsf{M}_{\boldsymbol{t}}\vdash)$ and $(\mathsf{b}_{\boldsymbol{t}}\vdash)$. By "half" of the equivalence relation we define an *ordering* on the equivalence classes as $[\mathcal{A}] \leq [\mathcal{B}]$ if $\boldsymbol{t} \vdash \mathcal{A} \to \mathcal{B}$. From the proof that the "whole" relation is an equivalence relation, it follows that $\leq$ is a partial order. (Instead of $[\mathcal{A}], [\mathcal{B}], [\mathcal{C}], \ldots$ we might use the usual algebraic notation $a, b, c, \ldots$ in quotient algebras, when it seems more convenient and there is no danger of confusion.)

QUOTIENT ALGEBRA OF $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$. The *quotient algebra* of $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$ is obtained by dividing out the *word algebra* of $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$ modulo the equivalence relation, and by defining operations on the equivalence classes by the binary connectives $\to$ and $\circ$ as well as by adding the partial order $\leq$. Then the result is $\mathfrak{A}^{\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}}_{\mathfrak{L}} = \langle A; \leq, \circ, \to, \boldsymbol{t}\rangle$, that is a *tonoid* satisfying (i)–(vii).

(i) $\langle A; \leq\rangle$ is a poset;

(ii) $\boldsymbol{t} \circ a = a$;

(iii) $a \circ b \leq c \ \Leftrightarrow\ a \leq b \to c$;

(iv) $a \leq b \wedge c \leq d \ \Rightarrow\ a \circ c \leq b \circ d$;

(v) $(a \circ b) \circ c \leq a \circ (b \circ c)$;

(vi) $(a \circ b) \circ c \leq b \circ (a \circ c)$;

(vii) $a \circ b \leq (a \circ b) \circ b$.

The *quotient algebra* of $\mathbf{LT}\text{-}\mathbf{W}^{\circ\boldsymbol{t}}_{\to}$ is alike except that (vii) does not hold.

We are not going to utilize directly $\mathfrak{A}^{\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}}_{\mathfrak{L}}$ or $\mathfrak{A}^{\mathbf{LT}\text{-}\mathbf{W}^{\circ\boldsymbol{t}}_{\to}}_{\mathfrak{L}}$, nevertheless, it is useful to keep (i)–(vi) and (vii) in mind when we define semantics for these logics. (For more explicit uses of algebraic methods see Dunn and Hardegree [2001].)

We hope to have demonstrated on the intensional fragments that it is *fruitful* — both conceptually and technically — *to refine* and *sharpen* various notions, such as proofs, sequent calculi, etc. As a result, *proof techniques* may have to (and sometimes can be) amended and improved — as is the case with the admissibility of the cut rule.

3.3 Semantics for $\mathbf{T}^{\circ\boldsymbol{t}}_{\rightarrow}$, $\mathbf{T}\text{-}\mathbf{W}^{\circ\boldsymbol{t}}_{\rightarrow}$, $\mathbf{T}\text{-}\mathbf{W}^{\boldsymbol{t}}_{\rightarrow}$ *and* $\mathbf{T}_{\rightarrow}$

There are various ways to interpret a logic. (For instance, in §2 we mentioned that **fde** has both a four-valued and a set theoretical semantics.) For $\mathbf{T}^{\circ\boldsymbol{t}}_{\rightarrow}$ we introduce an *operational semantics.* These kinds of semantics are often preferred in computer science, especially, when the logic interpreted is concerned with functions or programs. However, the naturally definable operations are not always compatible with the objects in the interpretation (as it happens, e.g., in the semantics of dynamic logic). For relevance logics operational semantics were introduced by Urquhart [1972] and also by Fine [1974]. The first semantics we give here is not based on a semi-lattice (as Urquhart's), nonetheless, a *binary operation* on the frame is central to the modeling of → and ∘. We incorporate into the operational setting some ideas from a semantics for residuated groupoids by Dunn [1993].

We follow the general pattern how *intensional logics* (in particular, modal logics) are interpreted in a set theoretical model. First, we define a *frame*, that is sometimes called a *model structure* or simply a *structure*, and then — relying on the frame — we define a *model* for the logic.

OPERATIONAL FRAMES FOR $\mathbf{T}^{\circ\boldsymbol{t}}_{\rightarrow}$. An *operational frame* $\mathfrak{F}$ for the logic $\mathbf{T}^{\circ\boldsymbol{t}}_{\rightarrow}$ is a quadruple $\langle U, \sqsubseteq, \circ, \iota\rangle$ where the following conditions are true.

(o1) $\sqsubseteq\, \subseteq U^2, \quad \circ\colon U^2 \longrightarrow U, \quad \iota \in U,$

(o2) $\alpha \sqsubseteq \alpha, \quad \alpha \sqsubseteq \beta \wedge \beta \sqsubseteq \gamma. \Rightarrow \alpha \sqsubseteq \gamma,$

(o3) $\alpha \sqsubseteq \beta \Rightarrow \alpha \circ \gamma \sqsubseteq \beta \circ \gamma,$

(o4) $\alpha = \iota \circ \alpha,$

(o5) $\alpha \circ (\beta \circ \gamma) \sqsubseteq (\alpha \circ \beta) \circ \gamma,$

(o6) $\alpha \circ (\beta \circ \gamma) \sqsubseteq (\beta \circ \alpha) \circ \gamma,$

(o7) $(\alpha \circ \beta) \circ \beta \sqsubseteq \alpha \circ \beta.$

Informally, the frame may be conceptualized similarly to the "possible worlds" (or relational, Kripke-style, set theoretical) semantics of modal logics. U is a preordered set of *situations* or *information states* and ι is a distinguished element of U — *the logical situation.* The operation ∘ is a binary operation that combines situations according to (o3)–(o7). (The symbol '∘' for this operation is chosen for easy remembrance.) For example, when a situation is fused with the logical one (from the right hand side) the same situation results.

In what follows we repeatedly use the notation $\wp(X)^{\uparrow}$ for the set of *upward closed* (*increasing*) subsets of (or *cones* on) a quasi-ordered set X. That is, given $\langle X, \sqsubseteq\rangle$, $Y \in \wp(X)^{\uparrow}$ when $Y \in \wp(X)$ and $x \in Y$ together with $x \sqsubseteq y$ implies $y \in Y$. Also, for any $Y \subseteq X$, $Y^{\uparrow}$ denotes the subset of X that contains all elements of Y as well as any element above (i.e., $Y^{\uparrow} = \{\, x\colon \exists y \in Y. y \sqsubseteq x \,\}$). Sometimes, $\{\, x \,\}^{\uparrow}$

may be abbreviated as $x^{\uparrow}$. *Directed (sub)sets* are a useful generalization of filters and ideals. Let $\langle X, \sqsubseteq \rangle$ be, again, a quasi-ordered set. $Y \subseteq X$ is *downward directed* when $x, y \in Y$ implies that $z \sqsubseteq x \wedge z \sqsubseteq y$ for some $z \in Y$.

MODELS FOR $\mathbf{T}_{\rightarrow}^{\circ \boldsymbol{t}}$. An *operational model* for $\mathbf{T}_{\rightarrow}^{\circ \boldsymbol{t}}$ is a pair $\langle \mathfrak{F}, v \rangle$ where $\mathfrak{F}$ is a frame for the logic and v is a function $v \colon \mathbb{P} \longrightarrow \wp(U)^{\uparrow}$.[28] The valuation is fixed for $\boldsymbol{t}$ by (o8) and it is extended to complex formulas by (o9) and (o10).

(o8) $v(\boldsymbol{t}) = \{\, \iota \,\}^{\uparrow}$,

(o9) $v(\mathcal{A} \circ \mathcal{B}) = \{\, \gamma \colon \exists \alpha \in v(\mathcal{A})\, \exists \beta \in v(\mathcal{B}).\, \alpha \circ \beta \sqsubseteq \gamma \,\}$,

(o10) $v(\mathcal{B} \rightarrow \mathcal{C}) = \{\, \alpha \colon \forall \gamma (\exists \beta \in v(\mathcal{B}).\, \alpha \circ \beta = \gamma \Rightarrow \gamma \in v(\mathcal{C})) \,\}$.

Note that the semantics is more than just operational because beyond the postulated binary operation we allowed a slight relational component "to sneak into" (o9). The valuation clauses may be stated (equivalently) using "truth" (1) and "falsity" (0), in a perhaps, more familiar form — and without formalizing the metalanguage. Let $v' \colon \mathbb{P} \times U \longrightarrow \{\, 0, 1 \,\}$ such that $v'(p, \alpha) = 1$ and $\alpha \sqsubseteq \beta$ implies $v'(p, \beta) = 1$. Then, (o8′)–(o10′) parallel (o8)–(o10).

(o8′) $v'(\boldsymbol{t}, \alpha) = 1$ if and only if $\iota \sqsubseteq \alpha$,

(o9′) $v'(\mathcal{A} \circ \mathcal{B}, \alpha) = 1$ if and only if $v'(\mathcal{A}, \beta) = 1$ and $v'(\mathcal{B}, \gamma) = 1$ for some β, γ such that $\beta \circ \gamma \sqsubseteq \alpha$ holds too,

(o10′) $v'(\mathcal{A} \rightarrow \mathcal{B}, \alpha) = 1$ if and only if for any β for which $v'(\mathcal{A}, \beta) = 1$ also $v'(\mathcal{B}, \alpha \circ \beta) = 1$.

The valuation v is restricted to a certain subset of the powerset of the set of situations, and it is possible to give a sufficient first-order condition (as we did in (o3)) on the pre-order and the binary operation that ensures that the extension of the valuation to compound formulas does not change v's co-domain.

LEMMA 11. (HEREDITY IN $\mathbf{T}_{\rightarrow}^{\circ \boldsymbol{t}}$) *For any formula* $\mathcal{A}$, $v(A) \in \wp(U)^{\uparrow}$.

THEOREM 12. (SOUNDNESS OF $\mathbf{T}_{\rightarrow}^{\circ \boldsymbol{t}}$) *If* $\mathcal{A}$ *is a theorem of* $\mathbf{T}_{\rightarrow}^{\circ \boldsymbol{t}}$ *then* $\iota \in v(\mathcal{A})$ *in any model for the logic.*

For the proof of completeness we define a canonical model (in a standard way) building it from pieces of the logic itself. Roughly, the idea is to take *theories*, or sets of formulas closed under implication with the additional property that any two formulas are implied by some formula contained in the theory.

CANONICAL FRAME FOR $\mathbf{T}_{\rightarrow}^{\circ \boldsymbol{t}}$. The *canonical frame* $\mathfrak{F}_{\mathfrak{c}}$ for $\mathbf{T}_{\rightarrow}^{\circ \boldsymbol{t}}$ is defined as $\mathfrak{F}_{\mathfrak{c}} = \langle U_{\mathfrak{c}}, \sqsubseteq_{\mathfrak{c}}, \circ_{\mathfrak{c}}, \iota_{\mathfrak{c}} \rangle$ where the members of the four-tuple are as follows.

(o11) $U_{\mathfrak{c}} = \{\, \alpha \colon \forall \mathcal{A}, \mathcal{B} (\mathcal{A} \in \alpha \wedge \vdash \mathcal{A} \rightarrow \mathcal{B}\, .\, \Rightarrow \mathcal{B} \in \alpha) \wedge \forall \mathcal{A}, \mathcal{B} (\mathcal{A}, \mathcal{B} \in \alpha \Rightarrow \exists \mathcal{C} \in \alpha .\, \vdash \mathcal{C} \rightarrow \mathcal{A} \wedge \vdash \mathcal{C} \rightarrow \mathcal{B}) \,\}$,

---

[28] Recall that $\mathbb{P}$ is the set of propositional variables of a logic. As a rule, we suppress indexing v (or its extension) with $\mathfrak{F}$ assuming that context prevents confusion, and that it is understood that without a frame, on which the model is defined, v is not defined at all.

(o12) $\sqsubseteq_\mathfrak{c}$ is the subset relation,

(o13) $\alpha \circ_\mathfrak{c} \beta = \{ \mathcal{C} \colon \exists \mathcal{A} \to \mathcal{B} \in \alpha \, \exists \mathcal{A} \in \beta . \vdash \mathcal{B} \to \mathcal{C} \}$,

(o14) $\iota_\mathfrak{c} = \{ \mathcal{A} \colon \vdash \boldsymbol{t} \to \mathcal{A} \}$.

The definition is fortunate in the sense that the canonical frame is in the class of $\mathbf{T}^{\circ \boldsymbol{t}}_{\to}$ frames.

LEMMA 13. (Canonical frame is frame) *The canonical frame* $\mathfrak{F}_\mathfrak{c}$ *of* $\mathbf{T}^{\circ \boldsymbol{t}}_{\to}$ *is a frame for* $\mathbf{T}^{\circ \boldsymbol{t}}_{\to}$*.*

Proof. There is a certain interest in proving this lemma, therefore, we include a few details. First we note that $\circ_\mathfrak{c}$ is easily proved to yield upward closed sets from similar ones. However, this is not enough since $U_\mathfrak{c}$ contains not arbitrary cones. Thus, let us assume that $\mathcal{C}_1, \mathcal{C}_2 \in \alpha \circ_\mathfrak{c} \beta$ where α, β are downward directed upward closed sets of formulas. By the definition of $\circ_\mathfrak{c}$ we obtain $\exists \mathcal{B}_1 \to \mathcal{A}_1 \in \alpha \, \exists \mathcal{B}_1 \in \beta . \vdash \mathcal{A}_1 \to \mathcal{C}_1$ as well as $\exists \mathcal{B}_2 \to \mathcal{A}_2 \in \alpha \, \exists \mathcal{B}_2 \in \beta . \vdash \mathcal{A}_2 \to \mathcal{C}_2$. The inductive hypothesis gives $\exists \mathcal{D} \in \alpha . \vdash \mathcal{D} \to (\mathcal{B}_1 \to \mathcal{A}_1) \wedge \vdash \mathcal{D} \to (\mathcal{B}_2 \to \mathcal{A}_2)$ and for β, $\exists \mathcal{E} \in \beta . \vdash \mathcal{E} \to \mathcal{B}_1 \wedge \vdash \mathcal{E} \to \mathcal{B}_2$. Using (A3) and (A2) we get $\vdash (\mathcal{D} \to (\mathcal{B}_1 \to \mathcal{C}_1)) \to (\mathcal{D} \to (\mathcal{E} \to \mathcal{C}_1))$. On the other hand, by two applications of (R1) to (A2) we have $\vdash (\mathcal{D} \to (\mathcal{B}_1 \to \mathcal{A}_1)) \to (\mathcal{D} \to (\mathcal{B}_1 \to \mathcal{C}_1))$. Then, by detachment from an instance of (A2) (or (A3)) together with the definition of $U_\mathfrak{c}$ we get $\vdash \mathcal{D} \to (\mathcal{E} \to \mathcal{C}_1)$. Alike steps give $\vdash \mathcal{D} \to (\mathcal{E} \to \mathcal{C}_2)$. These, respectively, amount to $\vdash \mathcal{D} \circ \mathcal{E} \to \mathcal{C}_1$ and $\vdash \mathcal{D} \circ \mathcal{E} \to \mathcal{C}_2$ by (R2). Also, $\mathcal{E} \to \mathcal{D} \circ \mathcal{E} \in \alpha$ because $\mathcal{D} \in \alpha$, hence, $\mathcal{D} \circ \mathcal{E} \in \alpha \circ_\mathfrak{c} \beta$. Then, by introducing conjunctions and an existential quantifier, we get $\exists \mathcal{G} \in \alpha \circ_\mathfrak{c} \beta . \vdash \mathcal{G} \to \mathcal{C}_1 \wedge \vdash \mathcal{G} \to \mathcal{C}_2$ as had to be shown. ■

To show that the logic is complete we need a suitable valuation on $\mathfrak{F}_\mathfrak{c}$.

Canonical valuation. The canonical valuation h is defined as

(o15) $h(\mathcal{A}) = \{ \alpha \colon \alpha \in U_\mathfrak{c} \wedge \mathcal{A} \in \alpha \}$.

THEOREM 14. (Completeness of $\mathbf{T}^{\circ \boldsymbol{t}}_{\to}$) *If* $\mathcal{A}$ *is not a theorem of* $\mathbf{T}^{\circ \boldsymbol{t}}_{\to}$ *then there is model for the logic such that the logical situation is not in* $v(\mathcal{A})$, *(i.e.,* $\nvdash \mathcal{A}$ *implies* $\iota \notin v(\mathcal{A})$ *in a model for* $\mathbf{T}^{\circ \boldsymbol{t}}_{\to}$*).*

First, one has to prove that the canonical valuation is indeed a valuation and this can be proven by induction on the structure of formulas. It is useful to note the step of the previous proof that we detailed as well as that any principal cone is a theory satisfying the additional condition of downward directedness. Second, h has to be proven not to validate any non-theorems.

The canonical model for $\mathbf{T}^{\circ \boldsymbol{t}}_{\to}$ can also be used to demonstrate that $\mathbf{T}_{\to}$ is not the implicational fragment of classical logic (as we hinted at on p. 736).

Adjusting the semantics for $\mathbf{T\text{-}W}^{\circ \boldsymbol{t}}_{\to}$ and $\mathbf{T\text{-}W}^{\boldsymbol{t}}_{\to}$. The connection between the combinatory axioms for B, B' and W, on the one hand, and (o5)–(o7), on the other, probably, does not require extra arguing. Indeed, as expected, by omitting (o7) we get a frame for $\mathbf{T\text{-}W}^{\circ\boldsymbol{t}}_{\rightarrow}$, and also for $\mathbf{T\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$. Based on the proofs of soundness (theorem 12), of the canonical frame lemma (lemma 13) and of completeness (theorem 14) soundness obtains for these logics and completeness obtains for $\mathbf{T\text{-}W}^{\circ\boldsymbol{t}}_{\rightarrow}$. However, a complication is caused by downward directedness in a proof of the canonical frame lemma for $\mathbf{T\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$.

The canonical frame for $\mathbf{T\text{-}W}^{\circ\boldsymbol{t}}_{\rightarrow}$ and $\mathbf{T\text{-}W}^{\boldsymbol{t}}_{\rightarrow}$ can be based on a set of situations $U_{\mathfrak{c}}$ that is defined as follows.

(o16) $U_{\mathfrak{c}} = \{\, \alpha \colon \forall \mathcal{A}, \mathcal{B} (\mathcal{A} \in \alpha \land \vdash \mathcal{A} \rightarrow \mathcal{B}. \ \Rightarrow \ \mathcal{B} \in \alpha) \,\}$.

Then *soundness* and *completeness* holds for both logics.

We noted earlier that adding $\circ$ as another binary connective gives a certain flexibility to go back and forth between provable sequents and provable formulas in a consecution calculus. Residuation is a plus in the interpretation too. There seems to be no straightforward way to prove lemma 13 for $\mathbf{T}^{\boldsymbol{t}}_{\rightarrow}$ without the canonical situations being downward directed objects, and there seems to be no way to prove that $\circ_{\mathfrak{c}}$ yields downward directed cones. Incidentally, there is a further subtlety here. It is not that unusual when the (standard) canonical frame of a nonclassical logic is not in the class of frames on which the logic is sound, because some frame condition is not satisfied by the (standard) canonical frame. Nevertheless, the canonical model is a model for the logic. (As an example take $\mathbf{G}$, the modal logic of provability, that is $\mathbf{K}$ together with the Löb axiom $\Box(\Box\mathcal{A} \supset \mathcal{A}) \supset \Box\mathcal{A}$. An analysis of the canonical frame of $\mathbf{G}$ is in Boolos [1995, Ch. 6].) In case of $\mathbf{T}^{\boldsymbol{t}}_{\rightarrow}$, the "canonical model" (so-called) in fact is *not* a model for the logic.[29] Certainly, one can claim that this shows — or at least strongly suggests — that the language of the pure implicational fragment of ticket entailment is too impoverished in its expressive power. However, another possible view is that the usual Hilbert-style proofs (what we called "ordinary proofs" at some point) do not provide sufficient information about the theorems they prove, and perhaps, we do not (yet) have a good understanding of the purely implicational theorems independently of some additional information.

The *semi-lattice semantics* for $\mathbf{T}_{\rightarrow}$ is due to Urquhart (see Anderson *et al.* [1992, §47.3]). It is related to a different formalization of the pure implicational fragment of ticket entailment. Therefore, before we outline the semi-lattice semantics we recall the *natural deduction* formulation of $\mathbf{T}_{\rightarrow}$. Actually, $\mathbf{T}_{\rightarrow}$ was first formulated as a Fitch-style natural deduction system by Anderson [1960]. $\mathbf{FT}_{\rightarrow}$ is defined like the Fitch-style systems for $\mathbf{E}$ and $\mathbf{R}$, but with some further restrictions. (For a detailed exposition of $\mathbf{FT}_{\rightarrow}$ together with the proof of its equivalence to $\mathbf{T}_{\rightarrow}$ see Anderson and Belnap [1975, 42–50].)

A *subproof* may contain wffs subscripted with finite subsets of $\mathbb{Z}^{+}$ and subproofs; so do proofs. Proofs are finite objects — just as in axiomatic calculi. The first item in a proof or subproof is a *hypothesis* which is a formula indexed with a singleton set that contains a positive integer which has not yet occurred in the proof and greater than all preceding numbers. Beyond this rule, which allows one to begin a proof or a subproof, there are four other rules.

---

[29] To be precise, the present author could not find a way to prove that to be a model.

(i) Any wff can be *repeated* within a subproof without changing its index set.

(ii) Any wff of the form $\mathcal{A} \to \mathcal{B}$ can be *reiterated* into a subproof with the same index set.

(iii) If $\mathcal{A} \to \mathcal{B}_\alpha$ and $\mathcal{A}_\beta$ occur in a subproof, and $\max(\alpha) \leq \max(\beta)$, then $\mathcal{B}_{\alpha\cup\beta}$ can be inserted into the subproof by $\to$ *elimination.*

(iv) If $\mathcal{A}_\alpha$ occurs within a subproof started with hypothesis $\mathcal{B}_{\{\,z\,\}}$ and $z \in \alpha$ ($z \in \mathbb{Z}^+$), then $\mathcal{B} \to \mathcal{A}_{\alpha-\{\,z\,\}}$ can be inserted into the preceding subproof by $\to$ *introduction.*

A formula $\mathcal{A}$ is *provable* in $\mathbf{FT}_\to$ if there is a proof of $\mathcal{A}_\emptyset$.

There are some simplifications of the above natural deduction system that do not change the set of theorems. For example, any instance of any theorem (already proven in another proof) can be inserted into a proof with the index $\emptyset$. (Note that the proviso "already proven" excludes circular proofs.) Requiring the hypotheses to be indexed by consecutive (increasing) numbers starting with 1 is a trivial, but convenient modification. Similarly, a *transitivity* rule (v) could be added to shorten proofs.

(v) If $\mathcal{A} \to \mathcal{B}_\alpha$ and $\mathcal{B} \to \mathcal{C}_\beta$ occur in a subproof then $\mathcal{A} \to \mathcal{C}_{\alpha\cup\beta}$ can be inserted into the subproof by *transitivity.*

It is easy to demonstrate that this rule is *derivable*, because the types of B and B' are, of course, provable, hence their appropriately relabeled proofs (or simply their appropriate instances indexed with $_\emptyset$) can be inserted into subproofs whenever needed. A more significant modification is to dispense with all the subproofs. Then the reiteration rule, which is the only rule essentially relying on subproofs, becomes superfluous due to the condition on the indices in (iii) — the so-called "ticket restriction." (The last observation is part of the proof of the equivalence of $\mathbf{FT}_\to$ to $\mathbf{T}_\to$ by Anderson and Belnap [1975, 45–46].) Omitting subproofs, however, forces a certain carefulness with $\to$ introductions. Namely, all these steps have to be made in decreasing order of the elements of the index set (after all hypotheses have been introduced).

It is rather easy to prove $(\mathcal{A} \to \mathcal{B}) \to ((\mathcal{B} \to \mathcal{A}) \to (\mathcal{A} \to \mathcal{B}))$ using $\mathbf{FT}_\to$ "from scratch," that is, without relying on the combinators we provided earlier. Finding a Hilbert-style proof without the help of the combinators is much more complicated. This illustrates that $\mathbf{FT}_\to$ has some advantages over $\mathbf{T}_\to$, especially, when it comes to practical proof search.[30] The index *sets*, while suffice for the purpose of defining an adequate proof system for the implicational fragment, make the combination of premises look simpler than sometimes desired.

---

[30] Natural deduction systems for **E** and **R** and some other related logics are overviewed in Anderson and Belnap [1975, §27.2]. Individual quantification for a wide range of natural deduction systems is introduced in Brady [1984]; propositional quantification is covered in Anderson *et al.* [1992, esp. §35.1].

SEMI-LATTICE SEMANTICS FOR $\mathbf{T}_\rightarrow$ (URQUHART, 1992). A *model* $\mathfrak{M}$ for $\mathbf{T}_\rightarrow$ is a five-tuple $\langle U, \vee, \bot, R, v\rangle$ where the members satisfy the following conditions.

(s1) $\vee\colon U^2 \longrightarrow U, \quad \bot \in U, \quad R \subseteq U^2, \quad v\colon \text{wff} \longrightarrow \wp(U)$,

(s2) $\langle U, \vee, \bot\rangle$ is a join semi-lattice with a bottom element,

(s3) R is a proto-order, that is, $R\alpha\beta \wedge R\beta\gamma. \Rightarrow R\alpha\gamma$,

(s4) $R\bot\alpha$,

(s5) $R\alpha\beta \Rightarrow R(\alpha \vee \gamma)(\beta \vee \gamma)$,

(s6) $v\colon \mathbb{P} \longrightarrow \wp(U)$,

(s7) $v(\mathcal{A} \rightarrow \mathcal{B}) = \{\, \alpha\colon \forall\beta(R\alpha\beta \wedge \beta \in v(\mathcal{A}). \Rightarrow \alpha \vee \beta \in v(\mathcal{B}))\,\}$.

A formula $\mathcal{A}$ is *valid* when in any model $\mathfrak{M}$ for $\mathbf{T}_\rightarrow$ $\bot \in v(\mathcal{A})$.

The soundness may be proven straightforwardly making use of the frame conditions. The completeness can be proven in two steps, first, constructing a canonical model from $\mathbf{FT}_\rightarrow$ and then establishing the equivalence of $\mathbf{FT}_\rightarrow$ and $\mathbf{T}_\rightarrow$ (cf. Anderson and Belnap [1975, §6]). For the purposes of the definition of the canonical model it is helpful to think of the index sets as *dependency sets*, that is, the indices referring to preceding formulas in the proof from which the wff has been derived. A theorem then is a wff that does not depend on any wff. $U_\mathfrak{c}$ is the set of *all finite sets of indexed wffs*, with $\emptyset$ the least element. $R_\mathfrak{c}\alpha\beta$ holds whenever $\forall\mathcal{A}_i \in \alpha\, \exists\mathcal{B}_j \in \beta.\ \max(i) \leq \max(j)$ or $\alpha = \emptyset$ (where $\max(i)$ is the greatest integer in i (if $i \neq \emptyset$), otherwise 0). Conditions (s1)–(s5) are satisfied, thus, it remains to define $v_\mathfrak{c}$. $\alpha \in v_\mathfrak{c}(\mathcal{A})$ iff there is a proof of $\mathcal{A}$ from α. (A proof of $\mathcal{A}$ from α is a sequence of indexed formulas ending with $\mathcal{A}$ such that (i) each indexed formula is in α, or a theorem of $\mathbf{T}_\rightarrow$ indexed with $_\emptyset$, or obtained by → elimination, and (ii) $\mathcal{A}$ depends on all members of α.)

The semi-lattice semantics and the Fitch-style natural deduction system are conducive to considering the connection between combinators and provable wffs more closely.

3.4 B, B′, W, I *combinators — a code for* $\mathbf{T}_\rightarrow$ *theorems*

We started this chapter by mentioning certain connections between the positive implicational logic (of Hilbert), a fragment of classical logic, and combinators as well as a similar connection between the implicational fragment of **R** and certain (other) combinators (cf. also Anderson *et al.* [1992, §71]). We pointed out that $\mathbf{T}_\rightarrow$ is related to I, B, B′, W (or I, B, B′, S′). Now we state the connection more precisely. A characterization of terms that can be abstracted using a I, B, B′, W *bracket abstraction* algorithm was first given by Helman [1977]. We recall a version of the notion of hereditary right maximal terms and a theorem that allows us to further reduce the set of inhabitants of the simple types of these terms. Then, we mention the result that $\mathbf{T}_\rightarrow$ may be formalized using the rule *condensed detachment*, which is reminiscent of, but not the same as, *modus ponens*.

The *simple type schemes* of S and K with *modus ponens* provide an axiomatization for the positive implicational fragment of classical logic (which coincides with the implicational fragment of intuitionistic logic, i.e., $\mathbf{H}_{\rightarrow}$). Traversing proofs may replace ordinary proofs in the positive implicational fragment of classical logic too. It is easy to see that if $\mathcal{A}$ is a theorem (of $\mathbf{H}_{\rightarrow}$) then there is a combinatory term built from S and K such that choosing appropriate instances of the axiom schemes and interpreting combinatory application as *modus ponens* the combinatory term represents (the proof of) the theorem $\mathcal{A}$. This observation can be made precise by using *Curry types* for combinators and for variables in a *type assignment system* (cf. e.g., Hindley [1997]). If the assignment $\tau\colon \mathcal{A}$ is provable, then the combinatory term τ has $\mathcal{A}$ as its simple type. A combinatory term has a *unique formation tree* and this tree corresponds to the natural deduction style proof tree of $\tau\colon \mathcal{A}$ in the type assignment system (assuming the minimally necessary consistent context). In the case of $\mathbf{H}_{\rightarrow}$ any theorem $\mathcal{A}$ has an *inhabitant*, a combinatory term τ, such that $\tau\colon \mathcal{A}$ is provable, moreover, any wff $\mathcal{B}$ for which $\tau\colon \mathcal{B}$ is provable, is a substitution instance of $\mathcal{A}$. In other words, every theorem is the *principal type* of an S, K combinator.

Of course, not all combinatory terms over $\{\, \mathsf{S}, \mathsf{K} \,\}$ are typable, which means that certain formation trees do not encode any proofs. To see what causes the problem one could first think of *Church types* for combinators and consider self-application. $\mathsf{K}^{\mathcal{A}\rightarrow(\mathcal{B}\rightarrow\mathcal{A})}\mathsf{K}^{\mathcal{A}\rightarrow(\mathcal{B}\rightarrow\mathcal{A})}$ clearly yields no implicational formula. However, KK (with Curry types) encodes a traversing proof of $\mathcal{A} \rightarrow (\mathcal{C} \rightarrow (\mathcal{B} \rightarrow \mathcal{C}))$, which is a — somewhat odd looking — theorem of classical logic (with $\supset$s instead of $\rightarrow$s). Nevertheless, the idea that some terms of the form $\tau\tau$ are not typable is fruitful for demonstrating that certain combinatory terms are not typable (with simple types). While combinators stand for sets of formulas, variables are limited to singletons (in consistent contexts), hence xx is simple typeless. Thus, for instance, the combinator $\mathsf{S}(\mathsf{SKK})(\mathsf{SKK})$ does not have a corresponding pure implicational formula. (The connection between the line of reasoning we sketched and this term is probably obvious, despite the fact that the combinator itself does not contain a subterm of the form $\tau\tau$.) In general, it is neither *necessary* nor *sufficient* for simple typelessness that a term is of the form $\tau\tau$. Famously, the combinator WI (i.e., M) does not have a simple type (though it has an intersection type). Another example of a combinatory term without a simple type is $\mathsf{W}(\mathsf{BBB}')$. The question if a term has a type is *decidable* and an algorithm can be defined to search for a simple type, moreover, for the principal one. In combinatory logic such an algorithm is called the *principal type search algorithm* (cf. Hindley [1997, 3E]).

The combinators S and K together have the remarkable property that they constitute a *combinatorially complete* basis. In other words, if there is a term τ built from (some of the) variables $x_1, \ldots, x_z$ (for some $z \in \mathbb{Z}^+$) then there is a combinator $\mathfrak{C}$ containing (solely) S and K such that $\mathfrak{C}x_1 \ldots x_z \rhd \tau$. In other words a *bracket abstraction* definable using these two combinators can simulate the abstraction of the $\lambda\mathsf{K}$-calculus.[31] Church's $\lambda\mathsf{I}$-abstraction limits the set of well-formed terms to those that do not contain any *vacuous abstraction.* A bracket abstraction definable using the combinators $\mathsf{I}, \mathsf{B}, \mathsf{C}, \mathsf{S}$, which correspond to $\mathbf{R}_{\rightarrow}$, is like the abstraction of the $\lambda\mathsf{I}$-calculus. In other words, if τ is a term that contains *all* variables $x_1, \ldots, x_z$ ($z \in \mathbb{Z}^+$), then there is an $\mathsf{I}, \mathsf{B}, \mathsf{C}, \mathsf{S}$ combinator $\mathfrak{C}$ with arity z such that $\mathfrak{C}x_1 \ldots x_z \rhd \tau$.

---

[31] The relationships between combinatory logics and λ-calculi (the detailed exploration of which

Although B' may be defined as CB, and so, obviously, includes *some permutation*, C is not definable from the basis $\{\,\mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{S}'\,\}$ (or equivalently, from $\{\,\mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{W}\,\}$). The set of terms that can be abstracted by an algorithm using these combinators was first characterized by Helman [1977]. We give a version of the definition of $\mathbf{HRM}_z$-terms that is suitable for our purpose here.

HEREDITARY RIGHT MAXIMAL TERMS. Let $\langle x_1, \ldots, x_z, \ldots\rangle$ be a denumerable sequence of variables. We define in parallel the *index* of a term (denoted by a subscript), and the *dependency set* of a term, denoted by a superscript as follows.

(t1) $\mathfrak{C}_0^{\{\,0\,\}}$ when $\mathfrak{C} \in \{\,\mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{W}\,\}$,

(t2) $y_i^{\{\,i\,\}}$ when y is x_i,

(t3) $M_i^{\alpha\cup\beta}$ when M is $N_j^{\alpha} P_k^{\beta}$ and $i = \max(j, k)$.

M_z^{α} is a *hereditary right maximal term* with variables up to and including x_z (i.e., $M_z^{\alpha} \in \mathbf{HRM}_z$, where $z \in \mathbb{Z}^+$) if (t4) and (t5) hold.

(t4) $k \geq j$ for all complex subterms $N_j^{\gamma} P_k^{\beta}$ of M_z^{α} if M is complex,

(t5) $\alpha = \{\,0, 1, \ldots, z\,\}$ or $\alpha = \{\,1, \ldots, z\,\}$.

$\mathbf{HRM}_z$-terms and $\mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{W}$ combinators are tightly connected. The set of $\mathbf{HRM}_z$-terms is *closed under reduction.* As a consequence, in the special case when a combinator (built from these combinators) is applied to (sufficiently many) variables subscripted with increasing positive integers, each reduction step yields an $\mathbf{HRM}_z$-term. If the term has a *normal form*, then that term is an $\mathbf{HRM}_z$-term too. On the other hand, given any $\mathbf{HRM}_z$-term τ, there is a combinatory term $\mathfrak{C}$ built from $\mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{W}$ that is a z-ary function term producing τ, that is, $\mathfrak{C}x_1 \ldots x_z \rhd \tau$.

The next lemma (except ‘principal’) follows from the existence of traversing proofs, for instance.

LEMMA 15. ($\mathbf{T}_{\rightarrow}$ THEOREMS AND $\mathbf{HRM}_z$-TERMS) *If $\mathcal{A}$ is a theorem of $\mathbf{T}_{\rightarrow}$ then there is an* $\mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{W}$ combinator $\mathfrak{C}$ *such that $\mathcal{A}$ is the* principal type scheme of $\mathfrak{C}$*; where the arity of $\mathfrak{C}^z$ ($z \in \mathbb{Z}^+$), is the number of $\mathcal{A}$'s antecedents. Further, $\mathfrak{C}x_1 \ldots x_z$ has a normal form which is an $\mathbf{HRM}_z$-term too.*

Of course, there are *infinitely* ($\aleph_0$) *many* $\mathbf{HRM}_z$-terms for any z. Furthermore, given a theorem $\mathcal{A}$ of $\mathbf{T}_{\rightarrow}$, $\mathfrak{C}$ in general, is *not unique.* The formula $(\mathcal{A} \rightarrow \mathcal{B}) \rightarrow$

---

is outside of the scope of this chapter), are more intricate than this broad statement might suggest. There are various relations between combinatory terms (e.g., weak reduction ($\rhd$), weak equality, strong reduction etc.), between λ-terms (e.g., β-reduction, $\beta\eta$-equality, etc.), and “the” bracket abstraction algorithm may be defined in more that one way.

$((\mathcal{B} \to \mathcal{A}) \to (\mathcal{A} \to \mathcal{B}))$, which was an example above, has over 20 inhabitants in normal form, each having the wff as its principal type scheme.

The following theorem provides a way to limit the number of potential inhabitants.[32] (Note that the choice of the particular combinatory basis is essential in this theorem.)

THEOREM 16. (I-FREE $\mathbf{HRM}_z$-TERMS (BIMBÓ, 2005)) *If $\mathcal{A}$ is a theorem of $\mathbf{T}_{\to}$ then it is either of the form $\mathcal{B} \to \mathcal{B}$ (that is, a type of I), or there is a combinatory term $\mathfrak{C}$ in normal form over the basis $\{\, \mathsf{B}, \mathsf{B}', \mathsf{W} \,\}$ such that $\mathcal{A}$ is a type of $\mathfrak{C}$.*

The idea behind limiting the set of inhabitants for a type is that this may lead to a solution of the decidability problem of the logic. Of course, it is decidable if a combinatory term τ can be assigned the type $\mathcal{A}$. (For instance, first one can find the principal type scheme (if any) of τ, then it is easy to check if $\mathcal{A}$ is a substitution instance.) However, the "converse" problem is in general not known to be decidable. That is, no algorithm is known that decides if an arbitrary simple type has an inhabitant among the combinators over a given combinatory basis. Therefore, limiting the set of potential inhabitants is clearly advantageous for solving the decidability problem of an implicational logic.

We devote the second part of this section to a brief outline of ***D**-completeness* — a proof theoretical property possessed by $\mathbf{T}_{\to}$ (formalized as (A1′)–(A4′) with rule ***D***, cf. theorem 17), however, not by $\mathbf{T\text{-}W}_{\to}$ (formalized similarly but without (A4′)).

Early on (in the first decades of the 20th century) classical logic was typically formalized with a rule of *substitution* included. This rule — which is not completely unlike the substitution rule in predicate logic or λ-calculus, and completely like substitution in combinatory logic — allows *uniform substitution* of wffs for propositional variables. (For instance, $(p_1 \to p_2) \to ((p_2 \to p_3) \to (p_1 \to p_3))[p_4 \to p_2/p_2, p_2/p_3, p_3/p_1]$ is $(p_3 \to (p_4 \to p_2)) \to (((p_4 \to p_2) \to p_2) \to (p_3 \to p_2))$.) Substitution preserves theoremhood or validity (except in a few exotic logics). Nowadays, usually axiom and rule *schemes* are used in the formalization of classical and nonclassical logics; this modification can be thought of as limiting substitution to the axioms and rules. (A perhaps useful analogy is the formulation of classical first-order logic without the rule of universal generalization, but with the axioms replaced by the sets of their universal closures.) Another way (that via schemes) "to get rid off" the rule of substitution in a purely implicational calculus is to build in a certain amount of substitution into the rule of *modus ponens.* Combining rules into new ones may lead to interesting new results (even if some resulting rules are perhaps not "very natural"). A relevant example is the modified implication introduction rule in $\mathbf{R}_{\to}$ which is helpful in showing decidability (see Dunn [1986]).

---

[32] For the proof of this theorem as well as for an application of the theorem see Bimbó [2005a].

The rule of *condensed detachment* allows just “the right amount” of substitution in the two premises to yield a principal type.[33] Then, an interesting question is if an implicational logic formalized by axiom schemes and *modus ponens* can be equivalently formalized by axioms that are instances of the schemes and condensed detachment, that is, if the axiomatization is ***D**-complete.* ‘Equivalently’, of course, means that the sets of theorems in the two calculi are the same. In a rather trivial sense any implicational logic, that has an axiomatization with *modus ponens*, can be made **D**-complete. (First, take the theorems as axiom schemes with detachment, then take the theorems as axioms with condensed detachment.) The equivalence of the sets of theorems is immediate, because the rule of condensed detachment is a *derivable rule* in any logic that includes *modus ponens* and the rule of substitution.

A slightly different notion of **D**-completeness is even more interesting. A logic is ***D**-complete* if it can be axiomatized by *finitely many* axioms and the rule of condensed detachment. (This definition attaches **D**-completeness to a logic as a set of theorems rather that a particular axiomatization. Thus an axiomatization of a logic may be **D**-incomplete, even though the logic is **D**-complete.) The logics defined by (A1) and (R1) and by (A2) and (R1) have been conjectured to be **D**-incomplete (in both senses).

RELETTERING WFFS. Formulas are finite and card($\mathbb{P}$) = $\aleph_0$. Therefore, it is immediate that any wff has denumerably many *substitution instances.* Moreover, any wff $\mathcal{A}$ has $\aleph_0$ instances produced by substitutions σ (where $\sigma\colon \mathbb{P} \longrightarrow \mathbb{P}$) that satisfy (rl) too.

(rl) $p_i \neq p_j \;\Rightarrow\; \sigma(p_i) \neq \sigma(p_j)$ whenever p_i, p_j occur in $\mathcal{A}$.

Given such a substitution σ, the formula $\sigma(\mathcal{A})$ that is the result of applying σ to $\mathcal{A}$ (where σ is recursively extended to complex wffs), is called a *relettering* or an *alphabetic variant* of the wff $\mathcal{A}$.

The “/”-notation, that we illustrated above, omits all p_i/p_is and all $\mathcal{A}/p_i$s if p_i does not occur in the affected wff; however, substitutions are formally total functions on $\mathbb{P}$. A relettering is just what the term says: σ may change the propositional letters of a wff to new ones, however, it may not identify them. (Reletterings may be thought of in more familiar terms as *permutations* on a (finite) subset of propositional variables; however, σ may reletter a wff without being injective on its whole domain.) Since formulas are finite objects, any two formulas have alphabetic variants that do not share any variables.

MOST GENERAL COMMON INSTANCE. Given two formulas $\mathcal{A}$ and $\mathcal{B}$ without a shared propositional variable, their *most general common instance* (*mgci*) results from substitutions σ_1 and σ_2 such that (c1) and (c2) hold.

---

[33] The rule of condensed detachment was invented by C. A. Meredith in the late 1950s — cf. Lemmon *et al.* [1969], Meredith and Prior [1963].

(c1) $\sigma_1(\mathcal{A}) = \sigma_2(\mathcal{B})$,

(c2) $\forall \sigma_3, \sigma_4(\sigma_3(\mathcal{A}) = \sigma_4(\mathcal{B}) \ \Rightarrow \exists \sigma_5, \sigma_6 . \, \sigma_5(\sigma_1(\mathcal{A})) = \sigma_3(\mathcal{A}) \wedge \sigma_6(\sigma_2(\mathcal{B})) = \sigma_4)$.

Without loss of generality, we may take any reletterings of $\mathcal{A}$ and $\mathcal{B}$ and any σ_1, σ_2 as long as they produce an *mgci*. Therefore, $\mathcal{A}$ and $\mathcal{B}$ either have *no mgci* or they have $\aleph_0$ *many mgci*s.

Informally, an *mgci* for a pair of wffs is a formula that is a substitution instance of both, and furthermore, any other wff that is a common substitution instance of the two formulas is a substitution instance of the *mgci*. From the definition it is obvious that any two *mgci*s are substitution instances of each other; then from the fact that no propositional variable is a substitution instance of an implication and that a substitution is a function it follows that *mgci*s are alphabetic variants of each other. (This outlines the essence of the *unification theorem*.)

Condensed detachment originally emerged as an abbreviation in writing out proofs, especially proofs demonstrating that fewer and shorter axioms suffice for some already known calculus. The condensed detachment of $\mathcal{B}$ from $\mathcal{A}$ is denoted by $\boldsymbol{D}\mathcal{A}\mathcal{B}$ in a prefix style notation, which is like $K\mathcal{A}\mathcal{B}$ for the conjunction and $C\mathcal{A}\mathcal{B}$ for the implication of $\mathcal{A}$ and $\mathcal{B}$. However, it should be noted that $\boldsymbol{D}$ is not a connective and $\boldsymbol{D}\mathcal{A}\mathcal{B}$ is not a wff. One way to actually, apply rule $\boldsymbol{D}$ is to generate a set of equations from the premises and solve them. This method became known as *unification* and has been used in theorem proving programs.

Condensed detachment may be formulated at least in two different ways. One places an emphasis on an *algorithm* that produces a resulting formula (or terminates with nil), the other formalizes the *idea* behind the rule that no unneeded substitutions are made in the premises of the detachment.

CONDENSED DETACHMENT, I. The *condensed detachment* of two formulas $\mathcal{A} \to \mathcal{B}$ and $\mathcal{C}$, which is either a wff or nil, is derived as follows.

(d1) Choose alphabetic variants of $\mathcal{A} \to \mathcal{B}$ and $\mathcal{C}$, let us say $\sigma_1(\mathcal{A} \to \mathcal{B})$ and $\sigma_2(\mathcal{C})$ that have no variables in common,

(d2) choose an *mgci* of $\sigma_1(\mathcal{A})$ and $\sigma_2(\mathcal{C})$ (if there is one), let us say, $\sigma_3(\mathcal{A})$,

(d3) output $\sigma_3(\mathcal{B})$ if (d1) is successful, otherwise, output nil.

The above description intended to "imitate" an algorithm (when it is understood that alphabetic variants and *mgci*s can be generated by algorithms), hence 'choose' may be read as 'generate the next' given some arbitrary but fixed ordering of variables. However, for theoretical (rather that implementational) purposes it is convenient to suppose that in (d1) and (d2) arbitrary alphabetic variants may be picked.

The next definition of condensed detachment is "nonalgorithmic," but it provides a better grasp of some properties of the resulting formula. ($\ell(\mathcal{A})$ is the length of the wff $\mathcal{A}$.)

CONDENSED DETACHMENT, II. The *condensed detachment* of two formulas $\mathcal{A} \to \mathcal{B}$ and $\mathcal{C}$ is a $\sigma_1(\mathcal{B})$ if (d4) is true for some σ_1, σ_2 (otherwise, it is nil), where σ_1 and σ_2 also satisfy (d5) and (d6).

(d4) $\sigma_1(\mathcal{A}) = \sigma_2(\mathcal{C})$,

(d5) $\forall \sigma_1', \sigma_2'. \sigma_1'(\mathcal{A}) = \sigma_2'(\mathcal{C}) \Rightarrow \ell(\sigma_1(\mathcal{B})) \leq \ell(\sigma_1'(\mathcal{B}))$,

(d6) $\forall \sigma_1', \sigma_2'. \sigma_1'(\mathcal{A}) = \sigma_2'(\mathcal{C}) \Rightarrow (\ell(\sigma_1'(\mathcal{B})) \leq \ell(\sigma_1(\mathcal{B})) \Rightarrow \exists \sigma_1''. \sigma_1''(\sigma_1(\mathcal{B})) = \sigma_1'(\mathcal{B}))$.

(d4) ensures that $\mathcal{A}$ and $\mathcal{C}$ have a common instance. (d5) states that $\sigma_1(\mathcal{B})$ is one of the shortest substitution instances obtained from a unifier of $\mathcal{A}$ and $\mathcal{C}$. Lastly, (d6) says that σ_1 keeps as many propositional variables in $\mathcal{B}$ distinct as possible.

THEOREM 17. ($\mathbf{T}_\to$ IS $\boldsymbol{D}$-COMPLETE (MEYER AND BUNDER, 1988)) *All theorems of* $\mathbf{T}_\to$ *are* principal types *of* $\{ \mathsf{I}, \mathsf{B}, \mathsf{B}', \mathsf{W} \}$ *combinators. That is, replacing* (A1)–(A4) *by*

(A1′) $p_0 \to p_0$

(A2′) $(p_0 \to p_1) \to ((p_2 \to p_0) \to (p_2 \to p_1))$

(A3′) $(p_0 \to p_1) \to ((p_1 \to p_2) \to (p_0 \to p_2))$

(A4′) $(p_0 \to (p_0 \to p_1)) \to (p_0 \to p_1)$

and taking rule $\boldsymbol{D}$ *instead of* (R1)*, all the theorems of* $\mathbf{T}_\to$ *are provable.*

An informal rendering of the role of the last three axioms is that (A2) and (A3) are needed to get to complex formulas from propositional variables, while (A4) can force the identification of propositional variables.

Other $\boldsymbol{D}$-complete logics include $\mathbf{R}_\to$ and $\mathbf{E}_\to$, as well as the implicational fragments of intuitionistic and classical logics. On the other hand, (A1)–(A3) with (R1) (an axiomatization of $\mathbf{T\text{-}W}_\to$) is not $\boldsymbol{D}$-complete, though $\mathbf{T\text{-}W}_\to$ is. It is possible to define infinitely many $\boldsymbol{D}$-complete logics, each strictly weaker that $\mathbf{T}_\to$ via essentially "hiding" a certain amount of contraction in instances of $\mathcal{A} \to \mathcal{A}$ without postulating (A4) itself (s. Megill and Bunder [1996]).

Incidentally, theorem 16 (about the I-free $\mathbf{HRM}_z$-terms) in a sense is the "opposite" of theorem 17. Instead of stating that for any theorem there is a combinator that has the formula as its principal type, that is, each theorem has its "own scape-goats," so to speak, theorem 16 practically omits one of the combinators in lieu of having certain theorems merely as substitution instances of a principal type of a combinator. The normal form restriction that is indispensable in the proof shrinks the set of inhabitants even further.

3.5 Positive fragment of ticket entailment: $\mathbf{LT}_+^{\circ t}$ *and* $\mathbf{T}_+^{\circ t}$

In **fde** there are no embedded implications, but **fde** contains two usual extensional connectives: conjunction and disjunction (cf. §2). In the intensional fragments these two connectives are not included (cf. §§3.1–3.3). Now we add the two *extensional connectives* to the intensional fragments. Compounding these parts can be motivated — beyond the inherent logical interest — by considering a wider range of inferences than before. We argued that a way of putting together premises (by $\circ$), which is more restricted than the "usual" one (by $\wedge$), is often reasonable and sometimes outright desirable. But there are other natural patterns of reasoning, some of which are more lax with respect to ordering and counting the premises, and perhaps, are better modeled using $\wedge$.

Conjunction and *disjunction* distribute over each other in classical logic and so do they in **fde**. To extend $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$ with $\wedge$ and $\vee$ one could think of adding the so-called "structure-free" rules for these connectives (cf. Belnap [1993]). These rules allow one to prove all the formulas that express the usual properties of $\wedge$ and $\vee$ except distributivity. The proof of distributivity requires some kind of contraction, permutation and thinning rules on the left in sequent calculi with a single right hand side. Given the connection between the *structural connective* ; and *fusion* $\circ$ the addition of exchange and weakening is not feasible; and the lack of associativity of ; is likely to cause further problems. Therefore, we adopt the approach that was invented for $\mathbf{R}^{\circ\boldsymbol{t}}_{+}$ (see $L\mathbf{R}^{+}$ in Dunn [1973] and Dunn [1986]) that postulates two structural connectives (cf. p. 737). The sequent calculus $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$**-W** — due to Giambrone (cf. Anderson *et al.* [1992, §67.2]) — is similar to our $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$ without ($\mathsf{W}\vdash$) (i.e., to $\mathbf{LT\text{-}W}^{\circ\boldsymbol{t}}_{+}$), but the two calculi are not exactly the same.

In the classical sequent calculus of Gentzen [1935] the comma , corresponds to $\wedge$ or $\vee$ depending on side. We add this structural connective to the previously introduced calculi (but keep them single right sided). Thus, in $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$ there are *two structural connectives* and the notion of stuctures is modified from the previous one (cf. p. 737).

Extensional and intensional structures. The *set of structures*, denoted by `str`, is the least set defined inductively as follows.

(i) If $\mathcal{A}$ is a wff then $\mathcal{A} \in \mathtt{str}_i$ and $\mathcal{A} \in \mathtt{str}_e$;

(ii) $\mathfrak{A} \in \mathtt{str} \wedge \mathfrak{B} \in \mathtt{str} \Rightarrow (\mathfrak{A};\mathfrak{B}) \in \mathtt{str}_i$;

(iii) $\forall^{z\in\mathbb{Z}^+}_{i\geq 2}\mathfrak{A}_i \in \mathtt{str} \Rightarrow \mathfrak{A}_1,\ldots,\mathfrak{A}_z \in \mathtt{str}_e$;

(iv) $\mathtt{str}_i \subseteq \mathtt{str} \wedge \mathtt{str}_e \subseteq \mathtt{str}$.

In words, formulas are the base set for both extensional and intensional structures; by (iii) , acts like a polyadic operation on structures. (ii) is exactly as before, that is, ; combines two structures into an intensional structure.

Sequent calculus $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$. A *consecution* is a structure followed by $\vdash$ and a formula. The *axiom* and *rule schemes* are (I), ($\to\vdash$), ($\vdash\to$), ($\circ\vdash$), ($\vdash\circ$), ($\mathsf{B}\vdash$), ($\mathsf{B}'\vdash$), ($\mathsf{W}\vdash$), ($\boldsymbol{t}\vdash$), ($\mathsf{b}_{\boldsymbol{t}}\vdash$) and ($\mathsf{M}_{\boldsymbol{t}}\vdash$) together with the following rules.

$$\frac{\mathfrak{A}[\mathcal{A}] \vdash \mathcal{C}}{\mathfrak{A}[\mathcal{A}\wedge\mathcal{B}] \vdash \mathcal{C}}\ {\scriptstyle \wedge_1\vdash} \qquad \frac{\mathfrak{A}[\mathcal{B}] \vdash \mathcal{C}}{\mathfrak{A}[\mathcal{A}\wedge\mathcal{B}] \vdash \mathcal{C}}\ {\scriptstyle \wedge_2\vdash} \qquad \frac{\mathfrak{A} \vdash \mathcal{A} \quad \mathfrak{A} \vdash \mathcal{B}}{\mathfrak{A} \vdash \mathcal{A}\wedge\mathcal{B}}\ {\scriptstyle \vdash\wedge}$$

$$\frac{\mathfrak{A}[\mathcal{A}] \vdash \mathcal{C} \quad \mathfrak{A}[\mathcal{B}] \vdash \mathcal{C}}{\mathfrak{A}[\mathcal{A}\vee\mathcal{B}] \vdash \mathcal{C}}\ {\scriptstyle \vee\vdash} \qquad \frac{\mathfrak{A} \vdash \mathcal{A}}{\mathfrak{A} \vdash \mathcal{A}\vee\mathcal{B}}\ {\scriptstyle \vdash\vee_1} \qquad \frac{\mathfrak{A} \vdash \mathcal{B}}{\mathfrak{A} \vdash \mathcal{A}\vee\mathcal{B}}\ {\scriptstyle \vdash\vee_2}$$

$$\frac{\mathfrak{A}[\mathfrak{B}] \vdash \mathcal{A}}{\mathfrak{A}[\mathfrak{B}, \mathfrak{C}] \vdash \mathcal{A}}\ {\scriptstyle K_e\vdash} \qquad \frac{\mathfrak{A}[\mathfrak{B}, \mathfrak{C}] \vdash \mathcal{A}}{\mathfrak{A}[\mathfrak{C}, \mathfrak{B}] \vdash \mathcal{A}}\ {\scriptstyle C_e\vdash} \qquad \frac{\mathfrak{A}[\mathfrak{B}, \mathfrak{B}] \vdash \mathcal{A}}{\mathfrak{A}[\mathfrak{B}] \vdash \mathcal{A}}\ {\scriptstyle W_e\vdash}$$

Notions such as proofs, provable sequents, cut rules are defined as before. Note that this calculus (similarly to $\mathbf{LT}^{\circ\boldsymbol{t}}_{\to}$) has the property that all provable sequents correspond to provable formula(s). The next theorem is provable like theorem 6.

THEOREM 18. (ADMISSIBILITY OF CUT RULE IN $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$) *The cut rule is* admissible *in* $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$*.*

The previous Hilbert-style calculus can be extended with conjunction and disjunction as follows.

POSITIVE TICKET ENTAILMENT AXIOMATIZED. The *axiomatic calculus* $\mathbf{T}^{\circ\boldsymbol{t}}_{+}$ is defined as an expansion of $\mathbf{T}^{\circ\boldsymbol{t}}_{\to}$ by axioms (A7)–(A13) and a rule (R6).

(A7) $(\mathcal{A} \wedge \mathcal{B}) \to \mathcal{A}$

(A8) $(\mathcal{A} \wedge \mathcal{B}) \to \mathcal{B}$

(A9) $((\mathcal{C} \to \mathcal{A}) \wedge (\mathcal{C} \to \mathcal{B})) \to (\mathcal{C} \to (\mathcal{A} \wedge \mathcal{B}))$

(A10) $\mathcal{A} \to (\mathcal{A} \vee \mathcal{B})$

(A11) $\mathcal{B} \to (\mathcal{A} \vee \mathcal{B})$

(A12) $((\mathcal{A} \to \mathcal{C}) \wedge (\mathcal{B} \to \mathcal{C})) \to ((\mathcal{A} \vee \mathcal{B}) \to \mathcal{C})$

(A13) $(\mathcal{A} \wedge (\mathcal{B} \vee \mathcal{C})) \to ((\mathcal{A} \wedge \mathcal{B}) \vee (\mathcal{A} \wedge \mathcal{C}))$

(R6) $\mathcal{A}, \mathcal{B} \ \Rightarrow\ \mathcal{A} \wedge \mathcal{B}$

Proofs can be defined as ordinary proofs were defined above — with the incorporation of the new axioms and rules; a *theorem* is a wff in a proof. Incidentally, it is interesting to observe that (R6) is a derived rule in $\mathbf{T}^{\circ\boldsymbol{t}}_{+}$ for theorems $\mathcal{A}, \mathcal{B}$.

THEOREM 19. (THEOREMS IN $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$ AND $\mathbf{T}^{\circ\boldsymbol{t}}_{+}$) *If* $\boldsymbol{t} \vdash \mathcal{A}$ *is provable in* $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$ *then* $\mathcal{A}$ *is a theorem of* $\mathbf{T}^{\circ\boldsymbol{t}}_{+}$*, and vice versa.*

We define an equivalence relation between formulas $\mathcal{A}$ and $\mathcal{B}$ to hold if and only if $\boldsymbol{t} \vdash \mathcal{A} \to \mathcal{B} \wedge \boldsymbol{t} \vdash \mathcal{B} \to \mathcal{A}$.

$\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$ ALGEBRAIZED. The *Lindenbaum algebra* of $\mathbf{LT}^{\circ\boldsymbol{t}}_{+}$ is $\mathfrak{A}_{\mathfrak{L}}^{\mathbf{LT}^{\circ\boldsymbol{t}}_{+}} = \langle A; \wedge, \vee, \to, \circ, \boldsymbol{t}\rangle$ (of similarity type $\langle 2, 2, 2, 2, 0\rangle$), where the following quasi-inequations are true.

(i) $\langle A; \wedge, \vee\rangle$ is a distributive lattice;

(ii) $\boldsymbol{t} \circ a = a$;

(iii) $a \circ b \leq c$ iff $a \leq b \to c$;

(iv) $(a \to b) \wedge (c \to b) \leq (a \vee c) \to b$ and $a \circ b \leq a \circ (b \vee c)$;

(v) $(a \circ b) \circ c \leq a \circ (b \circ c)$;

(vi) $(a \circ b) \circ c \leq b \circ (a \circ c)$;

(vii) $a \circ b \leq (a \circ b) \circ b$.

Informally, (ii) states that $\boldsymbol{t}$'s equivalence class is a (lower and upper) left identity. (iii) is two quasi-inequations expressing that $\to$ is the right residual of $\circ$. ((iii) can be equivalently stated by a finite set of equations.) (iv) gives certain tonicity properties to $\to$ and $\circ$. The last three conditions express properties of $\circ$ which are noticeably reminiscent of the axioms of the combinators B, B' and W.

A well-known feature of intuitionistic logic is that $\mathcal{A} \vee \neg\mathcal{A}$ is not a theorem. On the other hand, if $\mathcal{A} \vee \mathcal{B}$ is an intuitionistic theorem then so is $\mathcal{A}$ or $\mathcal{B}$.[34] It is not difficult to see how this property can be proven in **LJ** or in the single right sided $\mathbf{LT}_+^{\circ\boldsymbol{t}}$.

Disjunction property for $\mathbf{T}_+^{\circ\boldsymbol{t}}$. *The* disjunction property *is true in* $\mathbf{LT}_+^{\circ\boldsymbol{t}}$*, that is,* $\boldsymbol{t} \vdash \mathcal{A} \vee \mathcal{B}$ *implies that either* $\boldsymbol{t} \vdash \mathcal{A}$ *or* $\boldsymbol{t} \vdash \mathcal{B}$*.*

The same concept in "algebraese" is the *primeness* of an algebra. More concretely, in $\mathfrak{A}_{\mathfrak{L}}^{\mathbf{LT}_+^{\circ\boldsymbol{t}}}$ the *pseudo-(in)equation* (viii) holds.

(viii) $\boldsymbol{t} \leq a \vee b \;\Rightarrow\; \boldsymbol{t} \leq a \vee \boldsymbol{t} \leq b.$

For the semantics we are about to introduce for $\mathbf{T}_+^{\circ\boldsymbol{t}}$ it is useful to note the following property that we call *conjunction property.*

Conjunction property for $\mathbf{T}_+^{\circ\boldsymbol{t}}$. *The* conjunction property *is true in* $\mathbf{LT}_+^{\circ\boldsymbol{t}}$*, that is,* $\boldsymbol{t} \vdash (\mathcal{A} \wedge \mathcal{B}) \to \boldsymbol{t}$ *implies that either* $\boldsymbol{t} \vdash \mathcal{A} \to \boldsymbol{t}$ *or* $\boldsymbol{t} \vdash \mathcal{B} \to \boldsymbol{t}$*.*

3.6 Dual gaggle semantics for positive ticket entailment

Classical propositional logic has an easy-to-understand two-valued semantics, the existence of which is the trade-off for its capturing minimal content. Intensional logics such as normal modal logics and intuitionistic logic have "Kripke-style" *relational* semantics.[35] Unquestionably, the invention of the possible worlds semantics enhanced the understanding of these logics, for instance, by producing theorems on soundness, completeness, fmp and decidability for many of them. The relatively straightforward intuitive rendition of the maximally consistent sets of formulas as "worlds" and their relations as "accessibility" made these nonclassical logics more widely accepted and known. In a sense it became expected for an intensional logic to have some sort of set theoretical semantics, which was a shift from the earlier trend when semantics were primarily thought of as matrices.

Operational semantics, for example the semantics of $\mathbf{T}_{\to}^{\circ\boldsymbol{t}}$ above, cannot be adapted without considerable complications (such as two sorts of possible worlds) to the positive relevance logics, let alone to their extensions with negation — see Fine [1974] and Anderson *et al.* [1992, §51]. *Relational semantics* weaken the assumptions a structure has to satisfy (without sacrificing soundness), thereby leading to simpler and more elegant frames. At the end of the 60s, Dunn defined a relational semantics for the semi-relevant logic **R**-mingle (**RM**), using a *binary* accessibility relation; for the complete details see Dunn [1976b] and [1976c]. **RM** is not wholly relevant because it contains theorems like $\mathcal{A} \to \mathcal{B} \vee \mathcal{B} \to \mathcal{A}$ and $\mathcal{A} \to (\mathcal{A} \to \mathcal{A})$; the last one is just a notch from the positive and negative paradoxes. Various positive relevance logics were equipped with relational semantics in Routley and Meyer [1972b] — see also §3.8.

---

[34] Of course, classically it is the other way around. That is, the first wff is a theorem, and the second metatheorem is false.

[35] For an account of the emergence of the Kripke semantics see e.g., Goldblatt [2003].

A novelty of the *ternary relational semantics* was to use a relation with one more argument place, but perhaps even more important was the modeling of *implicational* formulas. *Gaggle theory* further generalizes relational semantics for nonclassical logics (see Dunn [1991], [1993], [1995]). ('Gaggle' expands 'ggl' that is the abbreviation of 'generalized Galois logics'.) All the above mentioned semantics exploit some kind of *theories*, that roughly speaking correspond to *filters* (for logics that include at least $\wedge$, and possibly, also $\vee$). The connection between theories and filters was explicated in the wider context of "generalized Kripke semantics" that were introduced in Bimbó and Dunn [2002]. Furthermore, Bimbó and Dunn [2005] also showed (among other things) how to "dualize" such semantics in the *absence of conjunction* relying on — slightly extrapolated — principles of gaggle theory. We give a similar semantics for $\mathbf{T}_+^{\circ\boldsymbol{t}}$.

DUAL GAGGLE FRAME FOR $\mathbf{LT}_+^{\circ\boldsymbol{t}}$. A *dual gaggle frame* for $\mathbf{LT}_+^{\circ\boldsymbol{t}}$ is a four-tuple $\mathfrak{F} = \langle U, \sqsubseteq, \bar{R}, \iota \rangle$ where the elements satisfy the following conditions.

(g1) $\sqsubseteq\ \subseteq U^2, \quad \bar{R} \subseteq U^3, \quad \iota \in U,$

(g2) $\bar{R}\alpha\beta\gamma \wedge \alpha \notin \iota^{\uparrow}.\ \Rightarrow\ \gamma \sqsubseteq \beta,$

(g3) $\exists\beta.\,\bar{R}\beta\alpha\alpha \wedge \beta \notin \iota^{\uparrow},$

(g4) $\bar{R}\uparrow\uparrow\downarrow$, i.e.,
$\bar{R}\alpha\beta\gamma \wedge \alpha \sqsubseteq \alpha' \wedge \beta \sqsubseteq \beta' \wedge \gamma' \sqsubseteq \gamma.\ \Rightarrow\ \bar{R}\alpha'\beta'\gamma',$

(g5) $\alpha \sqsubseteq \alpha, \quad \alpha \sqsubseteq \beta \wedge \beta \sqsubseteq \gamma.\ \Rightarrow \alpha \sqsubseteq \gamma,$

(g6) $\bar{R}\alpha\beta\gamma\ \Rightarrow\ \exists\delta.\,\bar{R}\alpha\beta\delta \wedge \bar{R}\delta\beta\gamma,$

(g7) $\bar{R}\alpha\beta\delta \wedge \bar{R}\delta\gamma\varepsilon\ \Rightarrow\ \exists\delta.\,\bar{R}\beta\gamma\delta \wedge \bar{R}\alpha\delta\varepsilon,$

(g8) $\bar{R}\alpha\beta\delta \wedge \bar{R}\delta\gamma\varepsilon\ \Rightarrow\ \exists\delta.\,\bar{R}\alpha\gamma\delta \wedge \bar{R}\beta\delta\varepsilon.$

Informally, the set U may be thought of as a set of *situations*. The smaller the situation with respect to the $\sqsubseteq$ ordering, the less informative it is, which means that $\sqsubseteq$ can be conceived as an information ordering. ι is the largest *alogical situation* and $\bar{R}$ is a ternary *accessibility* relation between situations.

DUAL GAGGLE MODELS FOR $\mathbf{LT}_+^{\circ\boldsymbol{t}}$. A *dual gaggle model* for $\mathbf{LT}_+^{\circ\boldsymbol{t}}$ is a pair $\mathfrak{M} = \langle \mathfrak{F}, v \rangle$ where $\mathfrak{F}$ is a frame for the logic and v is a valuation function of type $v\colon \mathbb{P} \longrightarrow \wp(U)^{\uparrow}$, that is fixed for $\boldsymbol{t}$ and extended to compound formulas as follows.

(g9) $v(\boldsymbol{t}) = \{\,\iota\,\}^{\uparrow},$

(g10) $v(\mathcal{A} \wedge \mathcal{B}) = v(\mathcal{A}) \cup v(\mathcal{B}),$

(g11) $v(\mathcal{A} \vee \mathcal{B}) = v(\mathcal{A}) \cap v(\mathcal{B}),$

(g12) $v(\mathcal{A} \circ \mathcal{B}) = \{\,\gamma\colon \forall\alpha, \beta(\bar{R}\alpha\beta\gamma \wedge \alpha \notin v(\mathcal{A}).\ \Rightarrow\ \beta \in v(\mathcal{B}))\,\},$

(g13) $v(\mathcal{B} \to \mathcal{C}) = \{\,\alpha\colon \exists\beta \notin v(\mathcal{B})\,\exists\gamma \in v(\mathcal{C}).\,\bar{R}\alpha\beta\gamma\,\}.$

Although there is a certain dual character to this semantics, the propositions are unchangeably upward closed sets.

LEMMA 20. (HEREDITY IN $\mathbf{T}_+^{\circ \boldsymbol{t}}$) *For any formula $\mathcal{A}$, $v(\mathcal{A}) \in \wp(U)^{\uparrow}$.*

THEOREM 21. (SOUNDNESS OF $\mathbf{T}_+^{\circ \boldsymbol{t}}$) *If $\mathcal{A}$ is a theorem of $\mathbf{T}_+^{\circ \boldsymbol{t}}$ then $v(\mathcal{A}) \subseteq \{\iota\}^{\uparrow}$.*

Notice that in the operational semantics the inclusion points into the opposite direction. The reason is exactly that the situations here may be thought of as (ramifications of) the "duals" of the situations in the operational semantics. For the proof of completeness we define the canonical frame, as usual.

CANONICAL FRAME FOR $\mathbf{T}_+^{\circ \boldsymbol{t}}$. The *canonical frame* of $\mathbf{T}_+^{\circ \boldsymbol{t}}$ is $\mathfrak{F}_\mathfrak{c} = \langle U_\mathfrak{c}, \sqsubseteq_\mathfrak{c}, \bar{R}_\mathfrak{c}, \iota_c \rangle$ defined as follows.

(g14) $U_\mathfrak{c} = \{\alpha\colon \forall \mathcal{A},\mathcal{B}(\mathcal{A} \in \alpha \wedge \vdash \mathcal{B} \to \mathcal{A}\,.\ \Rightarrow\ \mathcal{B} \in \alpha) \wedge \forall \mathcal{A},\mathcal{B}(\mathcal{A},\mathcal{B} \in \alpha\ \Leftrightarrow\ \mathcal{A} \vee \mathcal{B} \in \alpha) \wedge \forall \mathcal{A} \wedge \mathcal{B} \in \alpha.\ \mathcal{A} \in \alpha \vee \mathcal{B} \in \alpha\}$,

(g15) $\sqsubseteq_\mathfrak{c}$ is $\subseteq$,

(g16) $\bar{R}_\mathfrak{c}\alpha\beta\gamma\ \Leftrightarrow\ \forall \mathcal{A},\mathcal{B}(\mathcal{A} \circ \mathcal{B} \in \gamma\ \Rightarrow .\ \mathcal{A} \in \alpha \vee \mathcal{B} \in \beta)$,

(g17) $\iota_\mathfrak{c} = \{\mathcal{A}\colon\ \vdash \mathcal{A} \to \boldsymbol{t}\}$.

In words, the clauses mean that the canonical situations are *prime co-theories*, that is sets of wffs such that if a conjunction is in the co-theory then at least one of the conjuncts is in the co-theory; otherwise, co-theories are downward closed with respect to entailment, and closed with respect to disjunction. (Co-theories are (approximately) prime ideals on $\mathfrak{A}_{\mathfrak{L}}^{\mathbf{T}_+^{\circ \boldsymbol{t}}}$.) $\subseteq$ needs no explanation. $\bar{R}$ is the accessibility relation on the frame, which captures alternative choices (or branching) along fused wffs. ι is the set of all co-theorems of the logic. Note that the greatest co-theorem is, at the same time, the least theorem, and to this extent ι is not fully *a*logical.

LEMMA 22. (CANONICAL FRAME IS FRAME) *The canonical frame of $\mathbf{T}_+^{\circ \boldsymbol{t}}$ is a frame for the logic $\mathbf{T}_+^{\circ \boldsymbol{t}}$.*

We do not go into the details of this proof here, however, we recommend to the reader working out the proof, because this lemma (we believe) provides deep insights into this as well as into some other set theoretical semantics.

Canonical valuation. The *canonical valuation* h is defined as

(g18) $h(\mathcal{A}) = \{\, \alpha \colon \alpha \in U_{\mathfrak{c}} \wedge \mathcal{A} \in \alpha \,\}$.

THEOREM 23. (Completeness of $\mathbf{T}_{+}^{\circ t}$) *If $\mathcal{A}$ is not a theorem of $\mathbf{T}_{+}^{\circ t}$ then there is a model for $\mathbf{T}_{+}^{\circ t}$ in which $v(\mathcal{A}) \nsubseteq v(\boldsymbol{t})$.*

The structure of the proof of this theorem is similar to that of the previous completeness proofs. First, one proves by structural induction that h is indeed a valuation on the canonical frame, and second, that formulas that are not provably equivalent have different sets of situations corresponding to them in the model.

Adding both $\wedge$ and $\vee$ at once is a somewhat arbitrary choice. Of course, **fde** contains these two connectives, so adding them together is a move toward combining **fde** and the purely intensional fragments. However, either $\wedge$ or $\vee$ can be easily omitted by dropping the three rules from $\mathbf{LT}_{+}^{\circ t}$ that involve the connective. An inspection of the proof of ET for $\mathbf{LT}_{+}^{\circ t}$ confirms that *cut* is unchangeably *admissible.* Accordingly, either the disjunction or the conjunction property remains true. We denote by $\mathbf{LT}_{\vee}^{\circ t}$ the consecution calculus obtained by omitting $(\wedge_1 \vdash)$, $(\wedge_2 \vdash)$ and $(\vdash \wedge)$.

Soundness and completeness for $\mathbf{LT}_{\vee}^{\circ t}$. *The logic $\mathbf{LT}_{\vee}^{\circ t}$ is* sound *for the semantics modified by dropping* (g10)*. Moreover, if $\boldsymbol{t} \vdash \mathcal{A}$ is not provable in $\mathbf{LT}_{\vee}^{\circ t}$ then $h(\mathcal{A}) \nsubseteq h(\boldsymbol{t})$ in the canonical model which is defined as before except that the last conjunct is dropped from the definition of $U_{\mathfrak{c}}$.*

We emphasized that the logic of ticket entailment is *not the same* as classical logic. Broadly following ideas (that were articulated by others beforehand) we explicated that $\supset$ is not an implication. Thus, it is useful to point out that — despite appearances — there seems to have been a tendency in relevance logics to preserve as much from classical logic as can be reasonably retained. A clear example of this is **R**, but in a sense **E** supports this observation too. One of the principles that has been only partially questioned by **E** (and not at all by **R**) is the permutation of antecedents of an implication. $(\mathcal{A} \to (\mathcal{B} \to \mathcal{C})) \to (\mathcal{B} \to (\mathcal{A} \to \mathcal{C}))$, is a theorem of **R** but not of **E** or **T**. In **E** $\mathcal{B}$ has to be an implication to be permuted into the first position, that is, $(\mathcal{A} \to ((\mathcal{B} \to \mathcal{D}) \to \mathcal{C})) \to ((\mathcal{B} \to \mathcal{D}) \to (\mathcal{A} \to \mathcal{C}))$ is a theorem of **E** but not of **T**.

If $\circ$ is in the language and $\to$ is its residual, then permutation of the antecedents of an implication amounts to commutativity of $\circ$. Then $\boldsymbol{t}$ is not only a left but also a right identity (as it is in **R**). Therefore, $\boldsymbol{t}$ is *the* unique *identity* of $\circ$. Commutativity of $\circ$ implies that $\to$ is not only the right but also the left residual of $\circ$, or *the* residual.

Thus, it is possible to introduce a second "implication like" connective into the logic of ticket entailment without $(\mathcal{A} \to \mathcal{B}) \to (\mathcal{B} \leftarrow \mathcal{A})$ and $(\mathcal{B} \leftarrow \mathcal{A}) \to (\mathcal{A} \to \mathcal{B})$ becoming theorems.

Consecution calculus for $\mathbf{LT}^{\circ t}_{\underleftarrow{+}}$. The logic of ticket entailment with a *left residual* connective $\mathbf{LT}^{\circ t}_{\underleftarrow{+}}$ is $\mathbf{LT}^{\circ t}_{+}$ with $(\leftarrow\vdash)$, $(\vdash\leftarrow)$ and $(\mathsf{bcb}_{\boldsymbol{t}}\vdash)$ added.

$$\frac{\mathfrak{A}\vdash\mathcal{A}\quad\mathfrak{B}[\mathcal{B}]\vdash\mathcal{C}}{\mathfrak{B}[\mathfrak{A};\mathcal{B}\leftarrow\mathcal{A}]\vdash\mathcal{C}}\ {\scriptstyle\leftarrow\vdash}\qquad\frac{\mathcal{A};\mathfrak{A}\vdash\mathcal{B}}{\mathfrak{A}\vdash\mathcal{B}\leftarrow\mathcal{A}}\ {\scriptstyle\vdash\leftarrow}\qquad\frac{\mathfrak{A}[\boldsymbol{t};\mathfrak{B};\mathfrak{C}]\vdash\mathcal{A}}{\mathfrak{A}[\mathfrak{B};(\boldsymbol{t};\mathfrak{C})]\vdash\mathcal{A}}\ {\scriptstyle\mathsf{bcb}_{\boldsymbol{t}}\vdash}$$

Given the proofs of the preceding cut theorems it is not difficult to establish that *cut is admissible* in $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$.

The logic of positive ticket entailment with left residual, $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$, is axiomatized by adding rules (R7) and (R8) to $\mathbf{T}^{\circ t}_{+}$.

(R7) $\vdash(\mathcal{A}\circ\mathcal{B})\to\mathcal{C}\ \Rightarrow\ \vdash\mathcal{B}\to(\mathcal{C}\leftarrow\mathcal{A})$

(R8) $\vdash\mathcal{B}\to(\mathcal{C}\leftarrow\mathcal{A})\ \Rightarrow\ \vdash(\mathcal{A}\circ\mathcal{B})\to\mathcal{C}$

The two calculi, $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$ and $\mathbf{LT}^{\circ t}_{\underleftarrow{+}}$, are "in parallel" as were their ←less fragments.

THEOREM 24. ($\mathbf{LT}^{\circ t}_{\underleftarrow{+}}$ and $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$ are essentially same) *$\mathcal{A}$ is provable in $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$ if and only if $\boldsymbol{t}\vdash\mathcal{A}$ is a provable consecution in $\mathbf{LT}^{\circ t}_{\underleftarrow{+}}$.*

The admissibility of cut combined with theorems 19 and 24 gives a syntactic proof of the conservativity of $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$ over $\mathbf{T}^{\circ t}_{+}$. A semantic proof is immediate from the following.

Dual gaggle semantics for $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$. A *dual gaggle frame* for $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$ is exactly like that for $\mathbf{T}^{\circ t}_{+}$. A *dual gaggle model* is as before with the addition of a valuation clause for complex wffs containing ←.

(g19) $v(\mathcal{C}\leftarrow\mathcal{A})=\{\,\beta\colon\exists\alpha,\gamma.\,\bar{R}\alpha\beta\gamma\wedge\alpha\notin v(\mathcal{A})\wedge\gamma\in v(\mathcal{C})\,\}$.

The canonical frame and model is defined as before. $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$ is *sound* and *complete* for the dual gaggle semantics augmented with (g19).

Incidentally, the introduction of the ← connective (in addition to →) also ensures that provable sequents correspond to provable formulas. Thus, finally, we consider the case when the fusion operation ∘ is omitted from $\mathbf{T}^{\circ t}_{\underleftarrow{+}}$. In the consecution calculus this simply means dropping the two rules $(\circ\vdash)$ and $(\vdash\circ)$, while to obtain an adequate axiomatization one has to add two new rules (R9) and (R10) instead of (R2)–(R3) and (R7)–(R8).

(R9) $\vdash\mathcal{A}\to(\mathcal{B}\to\mathcal{C})\ \Rightarrow\ \vdash\mathcal{B}\to(\mathcal{C}\leftarrow\mathcal{A})$

(R10) $\vdash\mathcal{B}\to(\mathcal{C}\leftarrow\mathcal{A})\ \Rightarrow\ \vdash\mathcal{A}\to(\mathcal{B}\to\mathcal{C})$

The two resulting logics are denoted by $\mathbf{LT}^{t}_{\underleftarrow{+}}$ and $\mathbf{T}^{t}_{\underleftarrow{+}}$, and they can be shown to be equivalent as before.

Canonical frame for $\mathbf{T}^{t}_{\underleftarrow{+}}$. The notions of a *frame* and of a *model* need not to be modified at all (except, perhaps, omitting (g12)). However, in the definition of the canonical frame the definition of the canonical accessibility relation should be replaced by (g20) which is as follows.

(g20) $R_{\mathfrak{c}}\alpha\beta\gamma \;\Leftrightarrow\; \forall\mathcal{A},\mathcal{B}(\mathcal{B}\in\gamma \Rightarrow .\,\mathcal{A}\to\mathcal{B}\in\alpha \lor \mathcal{A}\in\beta)$.

$\mathbf{T}^{\boldsymbol{t}}_{\overset{+}{\leftharpoondown}}$ is *sound* and *complete* with respect to this dual gaggle semantics.

3.7 Negation in ticket entailment: $\mathbf{T}^{\circ\boldsymbol{t}}$

In §1 we emphasized that certain classically valid formulas which include negation are implausible as principles for implication. (1), (2) and $(\mathcal{A}\land(\sim\mathcal{A}\lor\mathcal{B}))\to{\sim}{\sim}\mathcal{B}$ are such wffs. We proved one of them not to be an **fde** theorem and it is an easy exercise to disprove the other two using the **fde** tableaux. Clearly, $\sim$ is not "classical negation," since $\land$ and $\lor$ have all their usual negation free properties.

De Morgan negation. A unary operation $\sim$ is called a *De Morgan negation* when (n1) and (n2) hold.

(n1) $\mathcal{A}\to{\sim}\mathcal{B} \;\Rightarrow\; \mathcal{B}\to{\sim}\mathcal{A}$, (n2) ${\sim}{\sim}\mathcal{A}\to\mathcal{A}$.

If conjunction and disjunction are in the language of the logic (with their usual properties) then (n3)–(n6), the De Morgan laws also hold.

(n3) ${\sim}(\mathcal{A}\land\mathcal{B})\to({\sim}\mathcal{A}\lor{\sim}\mathcal{B})$, (n5) ${\sim}(\mathcal{A}\lor\mathcal{B})\to({\sim}\mathcal{A}\land{\sim}\mathcal{B})$,

(n4) $({\sim}\mathcal{A}\lor{\sim}\mathcal{B})\to{\sim}(\mathcal{A}\land\mathcal{B})$, (n6) $({\sim}\mathcal{A}\land{\sim}\mathcal{B})\to{\sim}(\mathcal{A}\lor\mathcal{B})$.

Given a (bounded) distributive lattice and an order inverting unary operation $'$ of period two (like De Morgan negation) (n7) and (n8) may be false.

(n7) $a\land a'\leq b$, (n8) $b\leq a\lor a'$.

An operation $\sim$ satisfying (n1)–(n2) as well as (n7′) and (n8′) is called *orthonegation*, and if $\land$ and $\lor$ distribute then it is sometimes labeled "classical negation."

(n7′) $\mathcal{A}\land{\sim}\mathcal{A}\to\mathcal{B}$, (n8′) $\mathcal{B}\to\mathcal{A}\lor{\sim}\mathcal{A}$.

Axiomatic calculus for $\mathbf{T}^{\circ\boldsymbol{t}}$. The *logic of ticket entailment* $\mathbf{T}^{\circ\boldsymbol{t}}$ is defined by axioms (A1)–(A4), (A7)–(A13) and rules (R1)–(R6) together with axioms (A14)–(A16).

(A14) $(\mathcal{A}\to{\sim}\mathcal{A})\to{\sim}\mathcal{A}$

(A15) $(\mathcal{A}\to{\sim}\mathcal{B})\to(\mathcal{B}\to{\sim}\mathcal{A})$

(A16) ${\sim}{\sim}\mathcal{A}\to\mathcal{A}$

3.8 *Routley–Meyer-style relational semantics for* $\mathbf{T}^{\circ t}$

We introduce a *Routley–Meyer-style semantics* for $\mathbf{T}^{\circ t}$. Semantics for various systems of relevance logics (including fragments and extensions) were introduced in a series of co-authored papers of Meyer and Routley [1972], [1973], [1972a], [1972b]. The semantics we give for $\mathbf{T}^{\circ t}$ utilizes the idea of modeling both binary intensional connectives via a *three-place* accessibility relation, that is original with Routley and Meyer; however, we do not follow them in all the details (or in terminology). Instead we make certain modifications in line with Dunn and Meyer [1997] and Dunn [2001].

RELATIONAL FRAME FOR $\mathbf{T}^{\circ t}$. A *relational frame* for the logic $\mathbf{T}^{\circ t}$ is $\mathfrak{F} = \langle U, \sqsubseteq, R, {}^*, I\rangle$, where conditions (r1)–(r10) hold.

(r1) ${}^*\colon U \longrightarrow U, \quad I \neq \emptyset, \quad I \subseteq U,$

(r2) $\sqsubseteq \subseteq U^2$ is a quasi-order,

(r3) $R \subseteq U^3$ and $R{\downarrow}{\downarrow}{\uparrow}$,

(r4) $R\alpha\beta\delta \wedge R\delta\gamma\eta \;\Rightarrow\; \exists\delta.\, R\beta\gamma\delta \wedge R\alpha\delta\eta,$

(r5) $R\alpha\beta\delta \wedge R\delta\gamma\eta \;\Rightarrow\; \exists\delta.\, R\alpha\gamma\delta \wedge R\beta\delta\eta,$

(r6) $R\alpha\beta\gamma \;\Rightarrow\; \exists\delta.\, R\alpha\beta\delta \wedge R\delta\beta\gamma,$

(r7) $\alpha^{**} = \alpha,$

(r8) $R\alpha\beta\gamma \;\Leftrightarrow\; R\alpha\gamma^*\beta^*,$

(r9) $R\beta\beta^*\beta,$

(r10) $\exists\iota \in I.\, R\iota\alpha\beta \;\Leftrightarrow\; \alpha \sqsubseteq \beta.$

Informally, U is a set of *situations* and I is a nonempty subset of situations, the set of *logical situations.* (These situations are "akin" to the situations of the operational semantics of $\mathbf{T}^{\circ t}_{\rightarrow}$ and "duals" of those that were introduced in the gaggle type semantics for the positive fragment $\mathbf{T}^{\circ t}_{+}$.) $\sqsubseteq$ is a pre-order on the situations, again, it may be viewed as a kind of *information order.* * is a *unary operation* of period two, which also interacts with the *accessibility relation* R that connects certain situations and additionally obeys conditions (r3)–(r6) and (r8)–(r10). Informally, the accessibility relation may be thought of as relaxing an operation with the help of the pre-order. (Cf. in particular, (r10) and (o4).)

This seems to be a good place to briefly comment on the ternary accessibility relation's "intuitive plausibility." The binary accessibility relation in the semantics of nonclassical logics, especially, in the semantics of **H** or **S4** and some of its extensions, such as **S4.2**, **S4.3**, easily gives rise to mental images of accumulation of perfect knowledge through time and that of connections of time and necessity. Of course, we should not overlook that *not all* modal logics (not even all normal modal logics) are so easy to picture. For instance, **KB** is sound and complete for the class of frames with a symmetric relation — but it looks like there are not that many merely symmetric relations lingering out there in the realm of "intuitive interpretations."

Technically, the motivation for the binary accessiblity relation in the *Kripke-style* semantics of normal modal logics is that the set of maximally consistent sets of wffs, in general, is not closed under the desired operation definable from the modal connectives. Certainly, the use of a ternary accessibility relation to model binary intensional operations (or as in gaggle theory, the use of a $z+1$-ary relation to model a z-ary intensional operation) is technically perfectly well-motivated. Moreover, it seems to us that gaggle theory provides more explanatory power by treating various nonclassical logics in a *uniform* way than what *ad hoc* semantics built for particular logics can garner.

The *three-place* accessibility relation can be informally explicated depending on the logic which is interpreted. For instance, Dunn and Meyer [1997], crediting P. Woodroof, introduce a way of looking at the ternary relation as indexed binary relations in the semantics of structurally free logics. The binary intensional connective modeled there is application in the sense of combinatory logic. Since combinators are functions (in other words, programs), this type of interpretation is in harmony with other indexed binary relations entering semantics. Although there are important technical differences between structurally free logics and dynamic logic, a natural way to think of the (so-called) "standard interpretation" of dynamic logic is in terms of indexed binary relations.

In the semantics that we are about to define, the accessibility relation can be viewed as an encapsulation. If $R\alpha\beta\gamma$ holds then γ envelopes the two situations α and β put (or "fused") together. In other words, if $R\alpha\beta\gamma$ is true, then γ contains at least as much information as the combination of the two theories α and β, by which, of course, we mean all the possible pairings of elements of α and that of β (in this order) plus their consequences. To phrase the same thought slightly differently, R should not hold of $\langle\alpha,\beta,\gamma\rangle$ if it is possible to choose some information bearers $\mathcal{A}$ and $\mathcal{B}$, respectively from α and from β, and obtain some information via putting them together that is not contained in γ. Having explained frames informally, we proceed to define models.

RELATIONAL MODELS FOR $\mathbf{T}^{\circ\boldsymbol{t}}$. A *relational model* for $\mathbf{T}^{\circ\boldsymbol{t}}$ is $\mathfrak{M} = \langle\mathfrak{F}, v\rangle$ where $\mathfrak{F}$ is a relational frame for the logic and v is a valuation function mapping propositional variables and $\boldsymbol{t}$ into upward closed subsets of U that is extended to compound formulas as follows.[36]

(r11) $\alpha \vDash p$ iff $\alpha \in v(p)$,

(r12) $\alpha \vDash \boldsymbol{t}$ iff $\alpha \in v(\boldsymbol{t})$ iff $\iota \sqsubseteq \alpha$ for some $\iota \in I$,

(r13) $\alpha \vDash \mathcal{A} \wedge \mathcal{B}$ iff $\alpha \vDash \mathcal{A}$ and $\alpha \vDash \mathcal{B}$,

(r14) $\alpha \vDash \mathcal{A} \vee \mathcal{B}$ iff $\alpha \vDash \mathcal{A}$ or $\alpha \vDash \mathcal{B}$,

(r15) $\alpha \vDash \mathcal{A} \rightarrow \mathcal{B}$ iff for any β and γ such that $R\alpha\beta\gamma$, $\beta \vDash \mathcal{A}$ implies $\gamma \vDash \mathcal{B}$,

(r16) $\alpha \vDash \mathcal{A} \circ \mathcal{B}$ iff for some $\beta \vDash \mathcal{A}$ and for some $\gamma \vDash \mathcal{B}$ also $R\beta\gamma\alpha$,

(r17) $\alpha \vDash {\sim}\mathcal{A}$ iff $\alpha^* \nvDash \mathcal{A}$.

LEMMA 25. (HEREDITY IN $\mathbf{T}^{\circ\boldsymbol{t}}$) *For any formula $\mathcal{A}$, if $\alpha \vDash \mathcal{A}$ and $\alpha \sqsubseteq \beta$ then also $\beta \vDash \mathcal{A}$.*

---

[36] We use $\vDash$ for the extension of v; of course, in each case $\vDash$ is specific to the frame and the valuation, however, we refrain from subscripting it, because context prevents confusion.

The proof of this lemma is by structural induction, in which appealing to the tonicity condition of R (in particular, of the first and third argument places) is central.

THEOREM 26. (SOUNDNESS OF $\mathbf{T}^{\circ t}$) *If $\mathcal{A}$ is a theorem of $\mathbf{T}^{\circ t}$, then $\forall\alpha \in I^{\uparrow}.\, \alpha \vDash \mathcal{A}$.*

To prove completeness we first define a special structure. There is a certain analogy here with normal modal logics, however, it is not full. In classical logic and in classically based modal logics the theories used in proving completeness are maximal and they contain exactly one of each pair of formulas $\mathcal{A}$ and $\neg\mathcal{A}$. The canonical model of $\mathbf{T}^{\circ t}$ will not be restricted to contain only "consistent" situations — despite the fact that $\mathbf{T}^{\circ t}$ has no "contradictions" ($\mathcal{A} \wedge {\sim}\mathcal{A}$) among its theorems.

This time we define the canonical frame in two stages via the *Lindenbaum algebra.* Mutual provability defines an equivalence relation on the set of wffs; furthermore, $\mathbf{T}^{\circ t}$ has the *replacement property* with respect to this relation. Thus, the connectives cannot distinguish between the elements of the equivalence classes, that is, the connectives lead to well-defined operations.[37]

LINDENBAUM ALGEBRA OF $\mathbf{T}^{\circ t}$. The *quotient algebra* of $\mathbf{T}^{\circ t}$ is $\mathfrak{A}_{\mathfrak{L}}^{\mathbf{T}^{\circ t}} = \langle A; \wedge, \vee, \circ, \rightarrow, {\sim}, \boldsymbol{t}\rangle$ (of similarity type $\langle 2, 2, 2, 2, 1, 0\rangle$) where (i)–(x) are true.

(i) $\langle A; \wedge, \vee\rangle$ is a distributive lattice;
(ii) $\boldsymbol{t} \circ a = a$;
(iii) $a \circ b \leq c$ iff $a \leq b \rightarrow c$;
(iv) $a \circ (b \vee c) = (a \circ b) \vee (a \circ c)$;
(v) $(a \circ b) \circ c \leq a \circ (b \circ c)$;
(vi) $(a \circ b) \circ c \leq b \circ (a \circ c)$;
(vii) $a \circ b \leq (a \circ b) \circ b$;
(viii) ${\sim}{\sim} a = a$;
(ix) $a \leq {\sim} b$ iff $b \leq {\sim} a$;
(x) $a \circ b \leq c$ iff $a \circ {\sim} c \leq {\sim} b$.

To fix the meaning of certain terms we define filters and prime filters.

FILTERS AND PRIME FILTERS. Let $\mathfrak{A} = \langle A; \wedge, \vee\rangle$ be a distributive lattice. F is *a filter* when (F_1) and (F_2) hold.

(F_1) $F \in \wp(A) - \{\emptyset\}$,
(F_2) $a \in F$ and $b \in F$ iff $a \wedge b \in F$.

The set of all filters of $\mathfrak{A}$ is denoted by $\mathfrak{F}$. P is *a prime filter* when (F_1)–(F_2) as well as (P_1)–(P_2) hold of P.

(P_1) $P \in \wp(A)$,
(P_2) $a \in P$ or $b \in P$ iff $a \vee b \in P$.

The set of all prime filters of $\mathfrak{A}$ so defined is denoted by $\mathfrak{P}$.

Theories as *deductively closed sets of wffs* and *filters* of the Lindenbaum algebra of the logic are closely related. A (prime) filter is really a (prime) theory when one pretends not to notice that certain formulas are literally distinct though they can be proven to imply each other.[38] With this thought in mind we can define the canonical structure straightforwardly.

---

[37] As earlier we use lower case letters to denote the equivalence classes; e.g., we write a for the equivalence class generated by $\mathcal{A}$, and $\boldsymbol{t}$ stands for its own equivalence class.

CANONICAL STRUCTURE FOR $\mathbf{T}^{\circ\boldsymbol{t}}$. The *canonical frame* for $\mathbf{T}^{\circ\boldsymbol{t}}$ is defined as $\mathfrak{F}_{\mathfrak{c}} = \langle U_{\mathfrak{c}}, \sqsubseteq_{\mathfrak{c}}, R_{\mathfrak{c}}, {}^{*\mathfrak{c}}, I_{\mathfrak{c}} \rangle$ where the elements of the quintuplet are defined as

(r18) $U_{\mathfrak{c}} = \mathcal{P}$,

(r19) $\sqsubseteq_{\mathfrak{c}} = \subseteq$,

(r20) $R_{\mathfrak{c}} = \{ \langle \alpha, \beta, \gamma \rangle \colon \forall a, b.\, a \in \alpha \Rightarrow .\, b \in \beta \Rightarrow a \circ b \in \gamma \}$,

(r21) $\alpha^{*\mathfrak{c}} = \{ a\colon \sim a \notin \alpha \}$,

(r22) $I_{\mathfrak{c}} = \{ \iota\colon \iota \in \mathcal{P} \wedge \boldsymbol{t} \in \iota \}$.

The completeness proof of $\mathbf{T}^{\circ\boldsymbol{t}}$ is of particular interest, since $\mathbf{T}^{\circ\boldsymbol{t}}$ now includes implication, as well as all of **fde**. Furthermore, simply adding or subtracting some frame conditions but retaining the skeleton of the whole semantics one can furnish an interpretation for other relevance logics such as **B**, **E** and **R**. (This is not true for **L** though, not even for its nonmodal fragment. To obtain an adequate semantics for **L** quite a few significant changes must be made to the present relational semantics — cf. p. 778.)

CANONICAL STRUCTURE IN CLASS. *The* canonical structure $\mathfrak{F}_{\mathfrak{c}}$ *of* $\mathbf{T}^{\circ\boldsymbol{t}}$ *defined by* (r18)–(r22) *is a* frame *for* $\mathbf{T}^{\circ\boldsymbol{t}}$.

To make the structure of the completeness proof transparent, it is useful to prove first the so-called *squeeze lemma*.[39] Note that if $R_{\mathfrak{c}}\alpha\beta\gamma$ then after narrowing α and β to smaller prime theories or just smaller sets of wffs w, v the triplet $\langle w, v, \gamma \rangle$ satisfies the defining condition in (r20). (Of course, the variables in (r20) range over $U_{\mathfrak{c}}$ and w and v may or may not be members of $U_{\mathfrak{c}}$. Thus, it is convenient to denote by $R'_{\mathfrak{c}}$ the relation defined by (r20) without the assumption that α, β and γ are in $U_{\mathfrak{c}}$.) That is, $R'_{\mathfrak{c}}wv\gamma$, and similarly, $R'_{\mathfrak{c}}\alpha\beta u$ for any $u \supseteq \gamma$. These observations imply that $R_{\mathfrak{c}} \downarrow\downarrow\uparrow$. The squeeze lemma shows that under certain conditions it is possible to extend theories in the first two argument places of $R'_{\mathfrak{c}}$ to *larger theories*.

LEMMA 27. (SQUEEZE LEMMA) *Let* $R'_{\mathfrak{c}}\alpha\beta\gamma$ *where* $\alpha, \beta \in \mathcal{F}$ *and* $\gamma \in \mathcal{P}$. *Then there are* $\alpha', \beta' \in \mathcal{P}$ *such that* $\alpha \sqsubseteq_{\mathfrak{c}} \alpha'$ *and* $\beta \sqsubseteq_{\mathfrak{c}} \beta'$ *as well as* $R_{\mathfrak{c}}\alpha'\beta'\gamma$.

---

[38] Of course, the scrupulous reader noticed the "cheating" long ago. There is a type level difference between theories and filters. But $\equiv$ determines a homomorphism from the wffs, and via composition with a valuation (with the set of equivalence classes as its domain) we simply get another valuation (with the set of wffs as its domain).

[39] "Squeezing" is used repeatedly in the proof of completeness, therefore, it is useful to isolate the claim into a separately proven lemma. Different proofs of versions of this lemma may be found, e.g., in Routley and Meyer [1973] (cf. "primeness lemma"), Dunn [1986, §4.7], Anderson *et al.* [1992, §48.3].

Proof. The proof is in two steps, first we construct the two extended theories, and then we show that these theories are prime. We give a detailed sketch of the proof because of the importance of the lemma.
1 Given α and β we define a set E of ordered pairs of theories as follows.

$$E = \{\, \langle \alpha_0, \beta_0 \rangle \colon \alpha \subseteq \alpha_0 \wedge \beta \subseteq \beta_0 \wedge R'_{\mathfrak{c}}\alpha_0\beta_0\gamma \wedge \alpha_0, \beta_0 \in \mathcal{F} \,\}$$

Note that E is not empty, at least $\langle \alpha, \beta \rangle$ is an element of E. Further, with $\subseteq$ on pairs defined pointwise, E is a partially ordered set. For any linearly ordered set of pairs $\langle \alpha_i, \beta_i \rangle_{i \in I}$ (where the index set $I \neq \emptyset$) there is a least upper bound in E, namely, $\bigcup_{i \in I} \langle \alpha_i, \beta_i \rangle$ with $\bigcup$ defined componentwise. Zorn's lemma guarantees that E has a maximal element, and we denote this element by $\langle \alpha', \beta' \rangle$. We note that by the definition of the set E, obviously, $R'_{\mathfrak{c}}\alpha'\beta'\gamma$ as well as $\alpha \subseteq \alpha'$ and $\beta \subseteq \beta'$. The construction of α' and β' is complete; it remains to show that $R_{\mathfrak{c}}\alpha'\beta'\gamma$ holds as well.
2 Let us assume that $a_1 \vee a_2 \in \alpha'$ and neither $a_1 \in \alpha'$ nor $a_2 \in \alpha'$. From the construction of $\langle \alpha', \beta' \rangle$ follows that $\exists c_1, c_2 \in \alpha'$ and $\exists b_1, b_2 \in \beta'$ such that $(a_1 \wedge c_1) \circ b_1 \notin \gamma$ and $(a_2 \wedge c_2) \circ b_2 \notin \gamma$. $\alpha', \beta' \in \mathcal{F}$ and so $c_1 \wedge c_2 = c \in \alpha'$ as well as $b_1 \wedge b_2 = b \in \beta'$. By the tonicity of $\circ$ and by $\gamma \in \mathcal{P}$, we have that $(a_1 \wedge c) \circ b \notin \gamma$ and $(a_2 \wedge c) \circ b \notin \gamma$, hence neither $((a_1 \wedge c) \circ b) \vee ((a_2 \wedge c) \circ b) \in \gamma$. Then $((a_1 \vee a_2) \wedge c) \circ b \notin \gamma$, which contradicts E's definition. By eliminating the assumption to the contrary, α' is prime. The argument is symmetric with respect to the argument places of $\circ$. ■

The squeeze lemma is useful in showing that (r4)–(r6) and (r10) are true on $\mathfrak{F}_{\mathfrak{c}}$ as well as in some of the steps in the proof of the next theorem.

The canonical valuation maps an element a of the Lindenbaum algebra into the set of all objects from $U_{\mathfrak{c}}$ that contain a.

CANONICAL VALUATION. The *canonical model* $\mathfrak{M}_{\mathfrak{c}}$ is $\langle \mathfrak{F}_{\mathfrak{c}}, h \rangle$ where h is defined as

(r23) $h(a) = \{\, \alpha \colon \alpha \in \mathcal{P} \wedge a \in \alpha \,\}$.

Of course, h has to be shown to preserve the operations of the Lindenbaum algebra; then it follows that h induces a valuation from the set of wffs into $\wp(U_{\mathfrak{c}})^{\uparrow}$.

THEOREM 28. (COMPLETENESS OF $\mathbf{T}^{\circ t}$) *If in any model of* $\mathbf{T}^{\circ t}$ $\iota \vDash \mathcal{A}$ *for all* $\iota \in I$, *then* $\mathcal{A}$ *is a theorem of* $\mathbf{T}^{\circ t}$.

After h has been proven inductively to be a homomorphism, h has to be shown to be an isomorphism. If $a \nleq b$ (i.e., $\nvdash \mathcal{A} \to \mathcal{B}$) then the propositions that represent $\mathcal{A}$ and $\mathcal{B}$, respectively, are distinct. In other words, there is an element α of $U_{\mathfrak{c}}$ such that α separates ha from hb. $a^{\uparrow}$ fails to contain b, because of the initial assumption, hence $a^{\uparrow}$ can be maximalized (using Zorn's lemma) to a filter which does not contain b either. The Lindenbaum algebra is a distributive lattice which implies that such a *relatively maximal* filter is prime.

In §3.1 we emphasized the difference between ordinary and traversing proofs, although they generate the same set of $\mathbf{T}_{\to}$ theorems. Certain details of the soundness and completeness proof show that it is possible to define a consequence relation for $\mathbf{T}^{\circ t}$ in line with the concept of ordinary proofs. Let $\mathbb{A}$ be a set of wffs, and let $\mathbb{A} \vdash \mathcal{B}$ hold whenever from any set of wffs closed under the rules of $\mathbf{T}^{\circ t}$ there is a proof of $\mathcal{B}$ with elements of $\mathbb{A}$ allowed into the proof. Then the following holds.

THEOREM 29. (STRONG SOUNDNESS AND COMPLETENESS OF $\mathbf{T}^{\circ t}$) *$\mathbb{A} \vdash \mathcal{B}$ if and only if $\alpha \vDash \mathcal{A}$ for all $\mathcal{A} \in \mathbb{A}$ implies $\alpha \vDash \mathcal{B}$.*

We stressed throughout the chapter proving completeness in two steps, namely, by proving that the canonical structure is a structure, and then showing that the canonical model does not validate wffs that are not theorems. This is clearly analogous to standard ways of proving completeness for normal modal logics (when such a completeness proof is possible). If a modal logic is sound and complete for a class of frames and the canonical frame is a member of that class then the logic is called *canonical*. In other words, such a logic is sound with an arbitrary valuation on its canonical frame. ('An arbitrary valuation', of course, means an arbitrary valuation v with $\mathrm{dom}(v) = \mathbb{P}$ extended standardly to all wffs.) The algebra of a normal modal logic contains a Boolean algebra "in the background," (i.e., it has a **BA** reduct), and so the co-domain of the valuation function is simply $\wp(W)$ (where W is the set of possible worlds). The algebra of the logics we introduced so far are not Boolean algebras (with additional operations), and the valuation functions were defined with co-domain $\wp(U)^{\uparrow}$. Thus, the following theorem can be seen as a generalization of canonicity in modal logics.

THEOREM 30. (CANONICITY OF $\mathbf{T}^{\circ t}$) *Let v be a valuation ($v\colon \mathbb{P} \longrightarrow \wp(U_{\mathfrak{c}})^{\uparrow}$) on the canonical frame that is extended to $\vDash$ according to* (r11)–(r17). *Then, $\vdash_{\mathbf{T}^{\circ t}} \mathcal{A}$ only if $\iota \vDash \mathcal{A}$ for any $\iota \in I_{\mathfrak{c}}$ in the model $\mathfrak{M}_{\mathfrak{c}} = \langle \mathfrak{F}_{\mathfrak{c}}, \vDash \rangle$.*

3.9 More on the ticket entailment logics

The previous §§2–3.8 explored some of the ticket entailment systems. We included informal motivations behind various concepts, beyond providing definitions, statements of results and some details of (some of) the proofs. Now we turn to recalling some further important results that concern the ticket entailment logics (and often also the other relevance logics of §4).

THEOREM 31. (DECIDABILITY OF $\mathbf{LT}_{+}^{\circ t}$-$\mathbf{W}$ (GIAMBRONE, 1985)) *The positive fragment of ticket entailment without contraction, $\mathbf{LT}_{+}^{\circ t}$-$\mathbf{W}$, is* decidable.[40]

Giambrone's sequent calculus $\mathbf{LT}_{+}^{\circ t}$-$\mathbf{W}$ contains a $(\boldsymbol{t}{-}\vdash)$ rule, that is the "converse" of $(\boldsymbol{t}\vdash)$, that is, the constant $\boldsymbol{t}$ can be deleted from an intensional structure of the form $(\boldsymbol{t};\mathfrak{A})$. The $(\boldsymbol{t}{-}\vdash)$ rule is useful in proving the equivalence of the consecution calculus to the Hilbert-style axiomatization when the only structural rules are $(\mathsf{B}\vdash)$ and $(\mathsf{B}'\vdash)$. Then, a different but equivalent consecution calculus without $\boldsymbol{t}$ can be defined. Further, extensional structures can be simplified into ones without embedded extensional structures and without repetitions of formulas. The decidability result is based on a backward proof search in the consecution calculus with a limit on the number of potential antecedents for a provable formula. (With a certain modification to the sequent calculus, that is, by allowing unrestricted permutation, Giambrone also proved the decidability of $\mathbf{R}_{+}^{\circ \boldsymbol{t}}$**-W**.)

---

[40] The calculus $\mathbf{LT}_{+}^{\circ t}$-$\mathbf{W}$ (which is different from $\mathbf{LT}$-$\mathbf{W}_{+}^{\circ t}$) is defined in Giambrone [1985] (cf. also Anderson *et al.* [1992, §67]), where the theorem is proven too.

THEOREM 32. (UNDECIDABILITY OF RELEVANCE LOGICS (URQUHART, 1984)) *The deducibility problem for logics between* $\mathbf{T\text{-}W}_{+}$ *and* **KR** *is* undecidable*, and if* $((\mathcal{A} \rightarrow \mathcal{B}) \wedge \mathcal{A} \wedge \boldsymbol{t}) \rightarrow \mathcal{B}$ *is a theorem of the logic then it is* undecidable.

Deducibility means that given a finite set of wffs with elements $\mathcal{A}_1, \ldots, \mathcal{A}_n$ and a formula $\mathcal{B}$ there is a derivation of $\mathcal{B}$ from the set of wffs, that is, $\mathcal{B}$ is in the closure of the set $\{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$ under the rules of the logic. The problem of decidability, as we have mentioned, is the question whether the set of theorems of a logic is recursive or not. The logic **KR** is obtained from **R** via the addition of $(\mathcal{A} \wedge {\sim}\mathcal{A}) \rightarrow \mathcal{B}$. (Accidentally, **KR** is not the same logic as the "classical relevant logic" **CR** in §4.) Although **KR** is hardly relevant, it is still not the same as classical logic.

The theorem includes as special cases the *undecidability* of **T**, **E** and **R** and their positive fragments. The contrast between $\mathbf{T\text{-}W}_{+}$ and $\mathbf{T}_{+}$ seems to suggest an informal conjecture that "the type of W causes undecidability." However, it should not be forgotten that $\mathbf{R}_{\underset{\rightarrow}{\sim}}$ and $\mathbf{E}_{\underset{\rightarrow}{\sim}}$ — just as their negationless fragments — are *decidable*. Moreover, **R** without distributivity, and the semi-relevant logic **RM** have been shown *decidable* by Meyer (see Meyer [1966] and Anderson and Belnap [1975, §29.3]). (**RM** is **R** with $\mathcal{A} \rightarrow (\mathcal{A} \rightarrow \mathcal{A})$ added.) The finite model property for **T**, **E** and **R** would imply decidability, therefore, they *lack* the *fmp*.

The question of the decidability of **E** was one of the open problems posed by Anderson in the early 60s, and the problem of the decidability of **T**, **E**, **R** and of their positive fragments remained open for more than two decades. Thus it should not be surprising that Urquhart's proof is rather complex; it essentially relies on concepts and results that concern algebraic modular lattices together with a careful analysis of proofs — both within the logic $\mathbf{T\text{-}W}_{+}$ (and its extensions) as well as in the proof of the undecidability of the restricted word problem for finite dimensional projective geometries. The proof can be found (beyond the paper cited above) in Anderson *et al.* [1992, §65], and a brief conceptual overview of the proof by Urquhart is included in Brady [2003, §11.5].

There is a certain *philosophical importance* to undecidability. Harrop noted (in the mid 60s) that no philosophically interesting propositional logics were known to be undecidable, and that "[i]t would be of interest to a number of people to see a natural propositional calculus which has not got the finite model property, or possibly even better, one which is undecidable." Given that Harrop himself constructed several propositional calculi that do not possess the fmp but are decidable, his remarks gain in their weight. The decidability of a logic, of course, implies a certain "simplicity," although this need not be the same, in the sense of complexity theory, as that of the classical propositional logic. Not only classical and intuitionistic logics are decidable, but so is a wide range of propositional modal logics. Thus, relevance logics provided the *first example* of propositional logics that are *well-motivated* and *undecidable*, as contrasted with various "artificial" systems that were constructed specifically to be undecidable.

Interpolation and decidability are not closely related properties of a logic, in general, however, the following theorem was originally proven using some of the same geometric constructions as the previous one. By interpolation we mean again "Craig interpolation," that is, if $\mathcal{A} \to \mathcal{B}$ is provable in a logic then there is a wff $\mathcal{C}$ in the common vocabulary of $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A} \to \mathcal{C}$ and $\mathcal{C} \to \mathcal{B}$ are theorems.

THEOREM 33. (INTERPOLATION IN RELEVANCE LOGICS (URQUHART)) Interpolation fails *for all logics between* $\mathbf{T}\text{-}\mathbf{W}_+$ *and* **KR** (*independently of whether the language of the logic includes* $\boldsymbol{t}$ *and* $\boldsymbol{f}$ *or not*).[41]

The complete proof (with constants) is in Urquhart [1993], and the elimination of constants is in Brady [2003, §11.6]. The proof is somewhat complicated in the sense that it puts together *model theoretic* and *proof theoretic* theorems with theorems of *projective geometry.*

Frames with a ternary accessibility relation that are suitable for **KR** connect with projective geometry because R can be interpreted as the *colinearity* relation. In projective geometry it is possible to define a multiplication operation of points on a line. This operation is associative or nonassociative depending on properties of the projective geometry. In particular, there are projective geometries in which *Desargues's theorem* is not true. (A simple planar case of this theorem is illustrated in figure 2 (see p. 771), where the two triangles are perspectival with respect to the center of perspectivity $\mathcal{P}$ and with respect to the line ℓ.) All projective spaces with at least 3 dimensions are Arguesian, but in some projective spaces the theorem is not true. If the projective space has at least four coordinate points then the multiplication operation can be proven associative. A subtlety is that the algebraic proof of the associativity of the multiplication operation relies only on the modularity of the coordinate points and it is independent of the modularity of the other points.

The proof starts with selecting *two finite projective spaces*: a two dimensional non-Arguesian space and a three dimensional one. Concretely, these are $\pi_N(9)$ which is of order 9, and $PG(3,9)$ which is determined by the Galois field of order 9. A third projective plane, denoted by $\pi(3)$, can be shown to be a subplane of the two previous geometries. The next step is to define **KR** *frames from the geometries.* A **KR** frame is like an **R** frame (cf. §4) except that additionally the ternary relation R is "totally symmetric," that is, $R\alpha\beta\gamma$ implies $R\alpha\gamma\beta$ and $R\gamma\beta\alpha$. Although on frames of relevance logics this condition looks weird, it is useful to

---

[41] Actually, Urquhart's original claim is somewhat stronger, because his proof establishes the first part of the result up to $L(\mathcal{F}_1, \mathcal{F}_2)$, which is a logic that consists of wffs that are valid on two concrete relational frames $\mathcal{F}_1$ and $\mathcal{F}_2$. (We briefly outline the frame construction below, however, the projective spaces from which they are defined are only described by one of their names.)

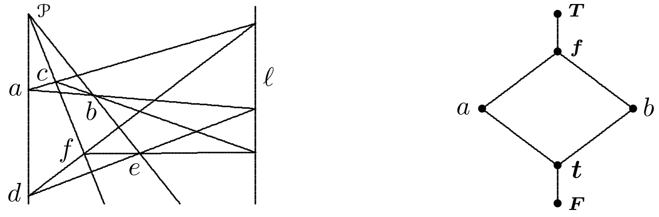


Figure 2. A planar case of Desargues's theorem and the lattice CRYSTAL

note that this condition is crucial to the proof of modularity for the connectives $\wedge$ and $\circ$. Colinearity — even if some special cases are excluded — is a relation which is invariant under the permutation of the arguments. Frames from $\pi_N(9)$ and $PG(3,9)$ are defined similarly by adding a point to the set of points of the geometries and turning colinearity into the accessibility relation. To be precise, R is colinearity of distinct points together with $R\alpha\alpha\alpha$ for all points α and $R\omega\alpha\alpha$ where ω is the added element. The frame obtained from $\pi_N(9)$ is further expanded by a new possible world to yield $\mathfrak{F}_1$; the frame defined from $PG(3,9)$ is $\mathfrak{F}_2$. The frame $\mathfrak{F}_0$ is defined from $\pi(3)$ by the addition of 15 points and several clauses (see Urquhart [1993, 465]).

$\mathfrak{F}_1, \mathfrak{F}_2$ are easily shown to be **KR** frames. The third frame is indeed a **KR** frame because it is possible to define maps from $\mathfrak{F}_1$ and $\mathfrak{F}_2$ that are onto $\mathfrak{F}_0$ and frame morphisms (i.e., surjective relational homomorphisms that are minimally faithful). These maps also ensure that certain theories defined on models on the two frames $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are extensions of a complete theory in their common vocabulary.

The next stage of the proof involves *encoding the multiplication operation* of points into logics. Since the frames ultimately come from projective geometries, the valuation can be defined using the points of the geometries. Informally, the content of the three theories T_0, T_1 and T_2 may be described as follows. T_0's vocabulary contains *three independent* points together with *three centers of perspectivity*; T_1 adds *three points* on a line, whereas T_2 adds a *fourth independent point* and *three centers of perspectivity.* The frame on which T_1 is defined is $\mathfrak{F}_1$, hence, T_1 proves that $\sim(A \cdot (B \cdot C) \leftrightarrow (A \cdot B) \cdot C)$. Since $PG(3,9)$ is Arguesian, the multiplication operation is associative, in other words, $T_1 \cup T_2$ proves that $A \cdot (B \cdot C) \leftrightarrow (A \cdot B) \cdot C$. **KR** contains **R** and so by rule (R6) and **KR**'s characteristic axiom $T_1 \cup T_2$ is inconsistent. This shows that the *joint consistency lemma* fails, because T_0 is a complete theory in a superset of the common vocabulary of T_1 and T_2 and the three theories themselves are consistent, however, $T_1 \cup T_2$ is inconsistent. (For classical first-order logic the joint consistency lemma was introduced, and its equivalence to the interpolation lemma was proved by A. Robinson.)

The last step of the proof requires a suitable *deduction theorem* to move from the failure of the joint consistency lemma to the failure of interpolation. This theorem ensures that given that there is a deduction of $\boldsymbol{f}$ from $T_1 \cup T_2$, there are formulas $\mathcal{A} \in T_1$ and $\mathcal{B} \in T_2$ such that $\mathcal{A} \rightarrow (\mathcal{B} \rightarrow \neg^{2l}\Box^k \boldsymbol{f})$ is a theorem, where $\neg^{2l}$ is $2l$-many $\neg$ s, $\Box\mathcal{D}$ is $\boldsymbol{t} \rightarrow \mathcal{D}$, and k is a positive integer indicating prefixing $\boldsymbol{t}$ k-times. Since $T_1 \cup T_2$ in the construction is inconsistent, by the deduction theorem, there are such formulas $\mathcal{A}$ and $\mathcal{B}$. Assuming that interpolation is true, $\mathcal{A} \rightarrow \mathcal{C}$ as well as $\mathcal{C} \rightarrow (\mathcal{B} \rightarrow \neg^{2l}\Box^k \boldsymbol{f})$ are theorems. $\mathcal{A} \in T_1$ implies $\mathcal{C} \in T_1$ because T_1 is a theory. However, then $\mathcal{C} \in T_0$, because $\mathcal{C}$ is in the common vocabulary, hence in T_2, because T_2 extends T_0. But from $\neg^{2l}\Box^k \boldsymbol{f} \in T_2$, $\boldsymbol{f} \in T_2$ follows, which contradicts T_2's consistency. (To ensure that the argument is applicable to the logics all the way from **KR** to $\mathbf{T\text{-}W}_+$, moreover, without constants further definitions within the logics and an analysis of proofs is necessary — see Anderson *et al.* [1992, §65], Urquhart [1993], Brady [2003, §11.6].)

Before we turn to listing various further properties of $\mathbf{T}^{\circ\boldsymbol{t}}$ we hasten to point out that the failure of interpolation *does not* imply the failure of variable sharing (with $\boldsymbol{t}$, or course, appropriately exempted). Indeed, in the logics between $\mathbf{T\text{-}W}_+$ and $\mathbf{R}^\circ$ *variable sharing* holds, that is, if $\mathcal{A} \rightarrow \mathcal{B}$ is a theorem then there is a propositional variable p occurring both in $\mathcal{A}$ and $\mathcal{B}$. A perhaps better way to look at the failure of interpolation is that these logics allow the formulation of theorems in which the antecedent relevantly implies the consequent, however, exactly because of the rigorous character of the implication the potential formulas built from the shared variables are either not implied by $\mathcal{A}$ or do not imply $\mathcal{B}$.

The first-degree fragment of **T** is **fde** and it is decidable, moreover, its *complexity* is co-NP (see §2). The decision procedure (if any) for the implication–conjunction fragment of **T** was established to be at least in EXPSPACE by Urquhart [1990]. A decision procedure for the pure implicational fragment of **R** was invented by Kripke (cf. Kripke [1959], Dunn [1986, §3.6]). In the same paper Urquhart showed that this algorithm for $\mathbf{R}^\circ_{\overset{\wedge}{\rightarrow}}$ is *primitive recursive in the Ackermann function.* The lower bound of any decision procedure for a logic between $\mathbf{T}_{\overset{\wedge}{\rightarrow}}$ and $\mathbf{R}_{\overset{\wedge}{\rightarrow}}$ was proven *not* to be *primitive recursive* by Urquhart [1999]. The proof is obtained from a space bounded version of the halting problem, where the space is limited by an "Ackermann kind" function. (Ackermann's original function was proven not to be primitive recursive — despite the fact that the function is easily definable by a handful of recursive equations. Some simple modifications of the function leave its characteristic feature, i.e., that it is not primitive recursive, intact.) The Ackermann kind functions are "fast groving" and they *dominate* every primitive recursive function. In other words, for every primitive recursive function there is a natural number such that the value of the Ackermann kind function is strictly greater for all greater numbers. To sum up, for instance, $\mathbf{R}_{\overset{\wedge}{\rightarrow}}$ is decidable, however, a decision procedure for this logic is not more or less infeasible, rather it is way out in the range of intractable.

The next notable results concerning the ticket entailment logics were obtained earlier. Logically valid inferences are often described by saying that the conclusion follows *inevitably* from the premises. Accordingly, in the context of relevance logics "the interesting modality" typically is necessity, which is taken to be characterized by principles reminiscent of the modal axioms of the minimal normal alethic modal logic, which is — coincidentally — also denoted by **T**. In $\mathbf{T}_{\rightarrow}$ there is *no formula in one variable* such that the corresponding modal principles would become provable — except, of course, the trivial case, when necessity of a wff is the wff itself. (For the proof see Anderson and Belnap [1975, §6].)

We mentioned earlier that in a certain formulation of the implicational fragment of classical logic the principal type scheme of K can be replaced by any wff $p \rightarrow (\mathcal{A} \rightarrow \mathcal{B})$ given that it is a two-valued theorem. Among the axioms of $\mathbf{T}_{\rightarrow}$ there is only $\mathcal{A} \rightarrow \mathcal{A}$ that allows a nonimplicational formula to be implied by or to imply a wff. If a logic does not have any theorem of the form $p \rightarrow (\mathcal{B} \rightarrow \mathcal{C})$ then it is said to have the *Ackermann property*; if a logic does not have a theorem that is like the converse of the previous wff, then it has the *converse Ackermann property.* $\mathbf{T}_{\rightarrow}$ does not have the converse Ackermann property, however, $\mathbf{T\text{-}W}_{\rightarrow}$ possesses it — cf. Anderson and Belnap [1975, 95–96].

Finally we turn to the relationship between **T** and classical logic, which is of particular philosophical interest. (The next theorem holds also for **T**, that is, when $\boldsymbol{t}$ and $\circ$ are omitted.)

THEOREM 34. (CLASSICAL LOGIC IN $\mathbf{T}^{\circ \boldsymbol{t}}$) *All the* theorems of classical logic *are theorems of* $\mathbf{T}^{\circ \boldsymbol{t}}$ *under a translation* τ*, where* τ *is defined as follows.*

$$\text{(i)}\ \ \tau(p) = p \ \text{ if } \ p \in \mathbb{P}; \quad \text{(ii)}\ \ \tau(\neg\mathcal{A}) = {\sim}\,\tau(\mathcal{A}); \quad \text{(iii)}\ \ \tau(\mathcal{A} \vee \mathcal{B}) = \tau(\mathcal{A}) \vee \tau(\mathcal{B}).$$

The proof of the previous theorem, probably, belongs to the "folklore" of relevance logics by now, but see Anderson and Belnap [1975, §24.1]. The proof of the next theorem, however, is not at all obvious.

THEOREM 35. (γ IN **T** (MEYER AND DUNN, 1969)) *The following rule is* admissible *in* **T** *as well as in* **E** *and* **R** (*and also in* $\mathbf{T}^{\circ \boldsymbol{t}}$).
(γ) $\vdash \mathcal{A},\ \vdash {\sim}\mathcal{A} \vee \mathcal{B} \ \Rightarrow \vdash \mathcal{B}.$

The rule (γ), which is also called "disjunctive syllogism," is obviously related — via matching the connectives as in the translation above — to the rule of *modus ponens* of classical logic (cf. Anderson and Belnap [1975, §16.1]). For a detailed exposition of how the admissibility of this rule became a question (originally for **E**) as well as for proofs of this theorem see Dunn [1986, §2.3] and Anderson *et al.* [1992, §25, also §28.2.4]. The first proof, which uses algebraic methods, was obtained by Meyer and Dunn [1969]. The rule is admissible, that is, for *theorems only.* A way to think of the admissibility of (γ) is that negation does not have a *fixed point* on any theorems of the logic. Negation in **T** is De Morgan negation and it is well-known that there are De Morgan lattices larger that any fixed finite size (and of size $\aleph_0$ too) in which some elements are identical to their negation (cf. also §2). Thus, the claim is far from trivial. Of course, the admissibility of (γ) also implies that for no wff $\mathcal{A}$ $\mathcal{A} \wedge {\sim}\mathcal{A}$ is a theorem of $\mathbf{T}^{\circ \boldsymbol{t}}$ (except if $\mathbf{T}^{\circ \boldsymbol{t}}$ is inconsistent, that is not). A yet another consequence of the admissibility of (γ) is that the rule $\vdash \mathcal{A} \wedge {\sim}\mathcal{A} \ \Rightarrow \vdash \mathcal{B}$ is admissible too, though $\nvdash \mathcal{A} \wedge {\sim}\mathcal{A} \rightarrow \mathcal{B}$, of course.

Lastly, we mention that classical negation can be *conservatively* added to $\mathbf{T}^{\circ t}$, and so $\mathbf{T}^{\circ t}$ can be *Gentzenized* in the framework of display logic (cf. Anderson *et al.* [1992, §62.5.4] and Brady [2003, §12.3]).

4 OTHER RELEVANCE LOGICS: **B**, **L**, **E** AND **R**

So far we mainly focused on the logic of ticket entailment and its fragments, although occasionally we also mentioned other logics. Admittedly, **T** is not one of the "most-written-about" relevance logics, which is one of the reasons we selected **T** as our main example. To use a metaphor, **T** has been kind of a younger sibling of the two older and better known relevance logics **E** and **R** — it's there, but not in the limelight.[42] Thus, it is reasonable to briefly introduce the last two logics as well as two others, **B** and **L**.

Even with four more logics — or rather, four more families of logics — introduced we neglect a whole range of topics some of which have direct philosophical applications. A *theory of relevant predication* has been worked out by Dunn (see Anderson *et al.* [1992, §74]), and systems of *relevant arithmetic* have been introduced and investigated by Meyer and Dunn (see Anderson *et al.* [1992, §§72–73]). (We list a few other themes — at the level of the detail of keywords — to hint at what we could not fit into a small chapter: the theory of implication; propositional quantification; individual quantification; meta-valuations; nonclassical metalogic; merge sequent calculi; semi-relevant systems; matrices; relevant theory of counterfactuals; relevant belief revision; depth relevance.)

To start with, we note that the implicational fragments of $\mathbf{B}, \mathbf{T\text{-}W}, \mathbf{T}, \mathbf{L}, \mathbf{E}$ and $\mathbf{R}$ are ordered by the following inclusions: $\mathrm{Th}(\mathbf{B}_{\to}) \subsetneq \mathrm{Th}(\mathbf{T\text{-}W}_{\to}) \subsetneq \mathrm{Th}(\mathbf{T}_{\to}) \subsetneq \mathrm{Th}(\mathbf{E}_{\to}) \subsetneq \mathrm{Th}(\mathbf{R}_{\to})$ and $\mathrm{Th}(\mathbf{B}_{\to}) \subsetneq \mathrm{Th}(\mathbf{T\text{-}W}_{\to}) \subsetneq \mathrm{Th}(\mathbf{L}_{\to}) \subsetneq \mathrm{Th}(\mathbf{R}_{\to})$ (where Th denotes the set of theorems); no other inclusions hold. **fde** is contained in the positive fragments of these logics except in **L**.

Basic relevance logic: **B**

There are various senses in which a logic can be "minimal." Although sometimes even $\mathcal{A} \to \mathcal{A}$ has been questioned as a reasonable principle for implication, (A1) is usually considered a wff that is acceptable to be an axiom. (Including (A1) in a logic is also helpful for defining a reasonable congruence relation on formulas.) $\mathbf{B}_{+}^{\circ t}$ was introduced by Meyer and Routley [1972]. $\mathbf{B}_{\to}$ consists of (A1) together with (R1), (R11) and (R12).

(R11) $\mathcal{A} \to \mathcal{B} \;\Rightarrow\; (\mathcal{B} \to \mathcal{C}) \to (\mathcal{A} \to \mathcal{C})$

(R12) $\mathcal{A} \to \mathcal{B} \;\Rightarrow\; (\mathcal{C} \to \mathcal{A}) \to (\mathcal{C} \to \mathcal{B})$

---

[42] A good example of this point is Meyer and Dunn [1969] that has **E** and **R** in its title, but includes the proof of the admissibility of (γ) for **T** too, as well as for some extensions of **E**.

The positive fragment of **B** is obtained by the addition of (A7)–(A13) and (R6) — while $\circ$ and $\boldsymbol{t}$ can be added as usual.

A *semantics* for $\mathbf{B}_+^{\circ t}$ can be obtained (cf. Routley and Meyer [1972b]) by omitting (r4)–(r6) and * from the definition of a frame for $\mathbf{T}^{\circ t}$ (together with all the conditions that involve *). (r17) obviously becomes superfluous in a model, so does (r21) in the definition of the canonical frame.

THEOREM 36. (Soundness, completeness (Meyer and Routley, 1972)) *$\mathcal{A}$ is a theorem of $\mathbf{B}_+^{\circ t}$ if and only if for any $\iota \in I^{\uparrow}$ $\iota \vDash \mathcal{A}$.*

A difference of $\mathbf{B}_+^{\circ t}$ from **fde** is that the former allows *embedded implications* to occur in provable formulas. However, $\mathbf{B}_+^{\circ t}$ does not postulate many features for the implication connective itself beyond self-identity. Another difference is, of course, that $\mathbf{B}_+^{\circ t}$ does not contain negation. The logic **B** (of Meyer and Routley [1972]) is defined by adding (A16) and (A17) as well as a rule form of contraposition (A15), that is, (R13).

(A17) $\mathcal{A} \vee {\sim}\mathcal{A}$

(R13) $\mathcal{A} \to {\sim}\mathcal{B} \;\Rightarrow\; \mathcal{B} \to {\sim}\mathcal{A}$

The corresponding semantic conditions for negation in the Routley–Meyer-style *relational semantics* for **B** are (r7), $\forall\iota \in I^{\uparrow}.\, \iota = \iota^*$ and $\exists\iota \in I^{\uparrow}.\, R\iota\alpha\beta \;\Rightarrow\; \exists\iota \in I^{\uparrow}.\, R\iota\beta^*\alpha^*$.

Meyer [2004] extended $\mathbf{B}_+$ with a Boolean negation yielding the logic **CB** ("classical **B**"), which is also shown there to be decidable.

Linear logic: L

Entailment is thought to include both *relevance* and *necessity* (see Anderson and Belnap [1962]). As a result, in relevance logics modality was investigated first of all as an operation *definable* from the implication connective of the logic itself (cf. the comment on the definability of $\Box$ in $\mathbf{T}_{\to}$ in §3.9), though $\Box$ also has been added as an undefined connective to **R** (cf. Meyer [1966] and Anderson *et al.* [1992, §61]). *Linear logic* was introduced by Girard [1987], who found a new use for a modality, so to speak. To put it in a slightly baffling way, the modalities in **L** signal that "resource consciousness" has been suspended with respect to the formula that is modalized. In other words, modalities allow the introduction of irrelevance, where relevance is understood in a more limited sense that in **R**, but in a more relaxed sense that in $\mathbf{B}^{\circ t}$.

Another difference of **L** from the logics considered so far in some detail (such as **fde**, $\mathbf{T}^{\circ t}$ and $\mathbf{B}^{\circ t}$) is the *lack of distributivity* of conjunction and disjunction. Distributivity of $\wedge$ and $\vee$ has not been questioned on the basis of concerns about relevance. However, a reason to consider relevance logics without distributivity existed nonetheless. Sometimes in a Hilbert-style axiomatization distributivity is an "extra" (independent) axiom in addition to the other axioms for $\wedge$ and $\vee$ (as

(A13) is above). Dropping distributivity allows a *simple sequent calculus* formalization without the need to introduce a second structural connective. Thus, **L** can be construed — undoubtedly, without faithfulness to history — as an extension of the *basic rigorous* relevance logic $\mathbf{BR}_+$ (see Anderson and Belnap [1975, 294]). $\mathbf{BR}_+$ adds to the distribution*less* part of $\mathbf{B}_+$ two further implicational axioms (A2) and (A18).

(A18) $(\mathcal{A} \to (\mathcal{B} \to \mathcal{C})) \to (\mathcal{B} \to (\mathcal{A} \to \mathcal{C}))$

This formula is the principal type scheme of C (cf. fn. 23). Linear logic then is a further extension of $\mathbf{BR}_+$ by $\circ$ and its dual $\oplus$ (that is sometimes called *fission*) as well as four propositional constants $\boldsymbol{t}, \boldsymbol{f}, \boldsymbol{T}, \boldsymbol{F}$ and negation $\sim$, but more crucially also by *two modalities* $\lozenge, \Box$, that are also duals of each other.[43]

One of the informal motivations of linear logic is the incorporation of a principle of consciousness about the use of resources (of computing devices). This relevance principle for implications excludes contraction but allows unlimited permutations yielding a notion of relevance stronger that in Church's weak calculus of implication, but weaker that in **T-W** or in **B** (cf. p. 774). The implicational fragment of linear logic can be axiomatized by the principal type schemes of the combinators B, C and I. Indeed, this logic was (independently) discovered, and named "B-C-I logic," by Meredith and Prior [1963] in the course of proof theoretical investigations of classical logic. The BCI axiomatization of $\mathbf{L}_\to$ exhibits an interesting *proof theoretic* phenomenon due to the lack of contraction.

LEMMA 37. ($\boldsymbol{D}$-INCOMPLETENESS OF AN AXIOMATIZATION OF $\mathbf{L}_\to$ (HINDLEY, 1997)) *The implicational logic defined by replacing the axiom schemes of the implicational fragment of linear logic* $\mathbf{L}_\to$ *by axioms, that is,* axiomatized *by* (A1′), (A2′) *and* (A18′) *together with condensed detachment is* $\boldsymbol{D}$-incomplete*.*

(A18′) $(p_2 \to (p_1 \to p_0)) \to (p_1 \to (p_2 \to p_0))$

However, it is possible to axiomatize the BCI logic differently. Theorem 16 emphasized that I has a special role in the BB′IW logic. Following theorem 17 we recalled a certain way of looking at the role of the type of W in that result. However, (A4) is not a theorem of BCI logic. Therefore, the "identification of variables" has to be handled differently if $\boldsymbol{D}$-completeness is to be attained.

LEMMA 38. ($\boldsymbol{D}$-COMPLETION OF BCI) BCI *is a* $\boldsymbol{D}$-complete logic *that may be axiomatized* (*equivalently to* (A1), (A2), (A18), (R1)) *by* (A1′), (A2′), (A3′), (A18′) *together with*

(A1″) $(p_1 \to (p_1 \to p_2)) \to (p_1 \to (p_1 \to p_2))$

and condensed detachment.

---

[43] The extensional connectives in **L** are called "additives," while the intensionals are "multiplicatives." The modalities are often labeled as "exponentials." Accordingly, the fragment without the modalities is called "exponential-free" or "multiplicative additive." (Cf. Avron [1988].)

Linear logic without modalities is defined by the axiom schemes and rules (A1)–(A2), (A18), (A7)–(A12), (A15)–(A16), and (A19)–(A21) as well as by the rules (R1)–(R8) where the additional axioms are

(A19) $(\boldsymbol{t} \to {\sim}\boldsymbol{f}) \wedge ({\sim}\boldsymbol{f} \to \boldsymbol{t})$

(A20) $(\mathcal{A} \to \boldsymbol{T}) \wedge (\boldsymbol{F} \to \mathcal{A})$

(A21) $({\sim}({\sim}\mathcal{A} \circ {\sim}\mathcal{B}) \to (\mathcal{A} \oplus \mathcal{B})) \wedge ((\mathcal{A} \oplus \mathcal{B}) \to {\sim}({\sim}\mathcal{A} \circ {\sim}\mathcal{B}))$

Obviously, some of the axioms can be viewed as definitions, and the set of connectives can be reduced without loss of the expressive power.

The similarity between the axioms and rules governing the modalities in **L** and in certain modal logics, in particular in **S4**, can easily lead to a misunderstanding as to the "nature" of these modalities. The axioms (A24)–(A26) and the rule (R14) below are reminiscent of the axioms (K), (T) and (4) and the rule of necessitation when $\Diamond$ is seen as $\Box$ — perhaps, with a little squinting. We note, however, that principles of modeling various connectives in *gaggle theory* make clear that the modality, originally denoted by !, is like a possibility operation, and its dual, originally denoted by ?, is like a necessity.

Therefore, to get full **L** as a Hilbert-system the following axioms and rule are added to the exponential-free fragment.

(A22) $\mathcal{A} \to (\Diamond\mathcal{B} \to \mathcal{A})$

(A23) $(\Diamond\mathcal{A} \to (\Diamond\mathcal{A} \to \mathcal{B})) \to (\Diamond\mathcal{A} \to \mathcal{B})$

(A24) $\Diamond(\mathcal{A} \to \mathcal{B}) \to (\Diamond\mathcal{A} \to \Diamond\mathcal{B})$

(A25) $\Diamond\mathcal{A} \to \mathcal{A}$

(A26) $\Diamond\mathcal{A} \to \Diamond\Diamond\mathcal{A}$

(R14) $\vdash \mathcal{A} \Rightarrow \vdash \Diamond\mathcal{A}$

(A22) and (A23) are "modalized" versions of the types of K and W. Axiom (A22) clearly shows the connection of the $\Diamond$ modality to irrelevance. As we mentioned (A24)–(A26) and (R14) (with $\Diamond$ replaced by $\Box$, and $\to$ replaced by $\supset$) suffice to define **S4** when added to classical logic. However, **L** is not classical and, for instance, $(\Diamond\mathcal{A} \wedge \Diamond\mathcal{B}) \to \Diamond(\mathcal{A} \wedge \mathcal{B})$ is not a theorem of **L**. Under the common intuitive interpretation of (alethic) modal logics (A25), of course, does not make a whole lot of sense — it states that "if $\mathcal{A}$ is possible, then $\mathcal{A}$ is the case." However, this does not prevent ! from being a $\Diamond$ sort of modality.

An *analytic tableaux* formulation of **L** is given in Meyer *et al.* [1995]. An alternative Gentzenization of linear logic was obtained by Belnap in the framework of *display logic* (cf. Anderson *et al.* [1992, §62]).

The absence of distributivity between the extensional connectives makes the definition of a sound and complete semantics more complicated that the Routley–Meyer semantics for $\mathbf{T}^{\circ\boldsymbol{t}}$ or $\mathbf{B}^{\circ\boldsymbol{t}}$. Girard [1987] defined an *operational semantics* for the multiplicative additive fragment. *Kripke models* with a ternary accessibility relation for the multiplicative additive fragment can be found in Allwein and Dunn [1993]. In terms of the Lindenbaum algebra of **L**, this semantics exploits maximally disjoint filter–ideal pairs with Galois connections, which were introduced as objects in a general lattice representation in Urquhart [1979]. A complication results, however, in the process of adding the intensional operations, because of the unprovability of the distributivity of conjunction and disjunction, the distributivity of fusion over $\vee$ cannot be proven either from the usual definition of these operations from the ternary accessibility relation that is used in the canonical model of $\mathbf{T}^{\circ t}$, for instance.

A less complex semantics was defined for *full linear logic* in Bimbó [2003]. The price for avoiding some of the complications is that when distributivity of $\wedge$ and $\vee$ is added to the logic, the resulting semantics slightly differs from the usual Routley–Meyer semantics for the relevance logic **B** and its extensions.

THEOREM 39. (UNDECIDABILITY OF **L** (LINCOLN *et al.*, 1992)) *Full propositional linear logic* **L** *is* undecidable.

The proof of Lincoln *et al.* [1992] is a *direct reduction* style proof, where the halting problem of a certain type of register machines is reduced to the problem of decidability of the logic. The computational model used is the and-branching two-register machine without zero tests. The halting problem for these machines is equivalent to the halting problem of Turing machines. The modality-free fragment of **L** ($\mathbf{L}\Diamond$) was shown decidable, in effect, in Meyer [1966] — the decidability of $\mathbf{L}\Diamond$ easily follows from the decidability of **R** without distribution, that is, from the decidability of *exponential-free linear logic with contraction.* For a computer science application of the latter result see Thistlewaite *et al.* [1988].

Lincoln *et al.* [1992] proved that the multiplicative additive fragment of linear logic is PSPACE-complete. Further, they also showed that the multiplicative fragment is in NP, whereas with weakening added the logic becomes NP-complete. The decision problem of *affine logic* (i.e., of the multiplicative modal fragment of **L** with weakening added) was solved by Kopylov [2001] via reducing affine logic to its "good" and then further to its "normal" fragment, which is proven decidable using a game on vectors that correspond to a certain type of fusion formulas. Affine logic is also known to have the fmp. Urquhart [2000] proved that the worst case complexity of Kopylov's decision procedure is not primitive recursive (rather primitive recursive in the Ackermann function — cf. section 3.9). He also stated that the decision procedure can be improved so that it had a double EXPSPACE upper bound.

Linear logic has various *computer science* motivations (such as modeling parallelism and proof-nets) which are to be found in Girard [1987]. A further motivation seems to be a translation of intuitionistic logic into **L**, in particular, a certain "decomposition" of intuitionistic implication. The translation function τ (for propositional **H**) is defined by the clauses (i)–(v).

(i) $\tau(p) = p$ whenever $p \in \mathbb{P}$;
(ii) $\tau(\mathcal{A} \wedge \mathcal{B}) = \tau(\mathcal{A}) \wedge \tau(\mathcal{B})$;
(iii) $\tau(\mathcal{A} \vee \mathcal{B}) = \Diamond\tau(\mathcal{A}) \vee \Diamond\tau(\mathcal{B})$;
(iv) $\tau(\mathcal{A} \to \mathcal{B}) = \Diamond\tau(\mathcal{A}) \to \tau(\mathcal{B})$;
(v) $\tau(\neg\mathcal{A}) = \Diamond\tau(\mathcal{A}) \to \boldsymbol{F}$.

Given the above translation if $\tau(\mathcal{A})$ is provable in linear logic then it is provable in **H**. The above decomposition of intuitionistic implication is directly linked to *coherent spaces* that provide an interpretation for intuitionistic logic. (This translation — not surprisingly — also lends itself to a translation of quantified λ-calculus.) Details of the connection to intuitionistic logic are in Girard [1987, §III.1 and §5.2]; the *quantum interpretation* of linear logic and **L**'s *philosophical significance* (via Aquinas's ideas) is outlined in Girard [2004].

Entailment: **E**

The discovery of a correspondence between certain implicational formulas and combinators goes back to Curry's and Church's work in the 30s and 40s (of the 20th c.) — cf. Curry and Feys [1958] and Hindley [1997]. Curry also used combinators in labels of structural rules in sequent calculi, for example, in Curry [1963]. To be able to further explicate the connection between the implicational logics and combinatory bases, we recall that there is a distinction between *proper* and not proper (*improper*) combinators in combinatory logic. Roughly speaking, a proper combinator applied to (sufficiently many) arguments gives a result in which only (some of) the arguments occur. The combinators K, I, B, B', C, W, M, S, and S' are all proper, and their axioms show the pattern we sketched. Not all combinators are proper, however. Famously, the *fixed point* combinator Y is improper. Moreover, the set of proper combinators is not closed under application. On the other hand, the set of simple typable combinators includes some improper combinators, but it does not include all proper combinators.

The implicational fragment of **E** can be defined by the principal types of I, B', S and $\mathsf{C}(\mathsf{B}'(\mathsf{B}'\mathsf{I}))$, that is, (A1), (A3), (A6) and (A27), together with rule (R1).

(A27) $(\mathcal{A} \to \mathcal{B}) \to (((\mathcal{A} \to \mathcal{B}) \to \mathcal{C}) \to \mathcal{C})$

is called *restricted assertion*, where "restricted" indicates that in assertion, that is, in $\mathcal{A} \to ((\mathcal{A} \to \mathcal{B}) \to \mathcal{B})$, the wff $\mathcal{A}$ has to be an implication. This restriction corresponds precisely to the fact that $\mathsf{C}(\mathsf{B}'(\mathsf{B}'\mathsf{I}))$ is improper. Assertion is the principal type of CI, but combinators in general cannot restrict arbitrary arguments to have complex or identical types, therefore, $\mathsf{B}'(\mathsf{B}'\mathsf{I})$ replaces I. Some of the other "restricted laws" that are theorems of $\mathbf{E}_{\to}$ (axiomatized as above) are principal types of other improper combinators. For instance, *restricted permutation* (rp) and *restricted conditional modus ponens* (rMP) are principal types of $\mathsf{C}(\mathsf{B}(\mathsf{CC})(\mathsf{B}'\mathsf{I}))$ and $\mathsf{B}(\mathsf{CC})(\mathsf{B}'\mathsf{I})$, respectively.

(rp) $(\mathcal{A} \to ((\mathcal{B} \to \mathcal{C}) \to \mathcal{D})) \to ((\mathcal{B} \to \mathcal{C}) \to (\mathcal{A} \to \mathcal{D}))$

(rMP) $(\mathcal{B} \to \mathcal{C}) \to ((\mathcal{A} \to ((\mathcal{B} \to \mathcal{C}) \to \mathcal{D})) \to (\mathcal{A} \to \mathcal{D}))$

These improper combinators do not have a commonly used single letter as abbreviation (like some proper ones have), but of course, we could name them as $1, 2$ and 3; then their axioms are $1xy \rhd y(\mathsf{B}'\mathsf{I}x)$, $2xyz \rhd xz(\mathsf{B}'\mathsf{I}y)$ and $3xyz \rhd yz(\mathsf{B}'\mathsf{I}x)$. The axioms clearly show both that these combinators are not proper as well as that their application leads out of the set of hereditary right maximal terms (cf. p. 750). The axioms also suggest that they themselves, or for instance $\{\,\mathsf{I}, \mathsf{B}', \mathsf{S}, 1\,\}$, cannot define C. (We used C in their original description to convey a certain familiarity.) Indeed, (A18) is a wff that is not a theorem of $\mathbf{E}_{\to}$.

The implicational fragment of $\mathbf{E}$ is *$\boldsymbol{D}$-complete*, which is a consequence of the $\boldsymbol{D}$-completeness of $\mathbf{T}_{\to}$ and $\mathbf{E}_{\to}$'s passing the identity test.

NECESSITY IN $\mathbf{E}_{\to}$ (ANDERSON AND BELNAP, 1962). There is a wff of $\mathbf{E}_{\to}$ in one variable such that (with that wff as a definition of $\Box$) the following wffs are *theorems* of $\mathbf{E}_{\to}$.

($\Box_1$) $(\mathcal{A} \to \mathcal{B}) \to (\Box\mathcal{A} \to \Box\mathcal{B})$

($\Box_2$) $(\mathcal{A} \to \mathcal{B}) \to \Box(\mathcal{A} \to \mathcal{B})$

($\Box_3$) $\Box\mathcal{A} \to \Box\Box\mathcal{A}$

From ($\Box_2$), and (R1), it is immediate that the rule, which is in normal modal logics called "necessitation" is *admissible.*

($\Box_4$) $\vdash \mathcal{A} \Rightarrow \vdash \Box\mathcal{A}$

The formula that defines $\Box\mathcal{A}$ is $(\mathcal{A} \to \mathcal{A}) \to \mathcal{A}$. (For further discussion of the role of necessity and of the provability of the above wffs, as well as some other modal formulas, see Anderson and Belnap [1975, §4.3].)

The implicational fragment can be extended with $\wedge$, $\vee$ via adding the axioms and rules (A7)–(A13) and (R6) as well as axiom (A28).

(A28) $(((\mathcal{A} \to \mathcal{A}) \wedge (\mathcal{B} \to \mathcal{B})) \to \mathcal{C}) \to \mathcal{C}$

This wff ensures that $(((\mathcal{A} \to \mathcal{A}) \to \mathcal{A}) \wedge ((\mathcal{B} \to \mathcal{B}) \to \mathcal{B})) \to (((\mathcal{A} \wedge \mathcal{B}) \to (\mathcal{A} \wedge \mathcal{B})) \to (\mathcal{A} \wedge \mathcal{B}))$ is a theorem, which — according to the definition of $\Box$ above — can be abbreviated simply as $(\Box\mathcal{A} \wedge \Box\mathcal{B}) \to \Box(\mathcal{A} \wedge \mathcal{B})$. Distributivity of $\Box$ over $\wedge$ is a desirable property of $\Box$ when $\Box$ is thought of as necessity.

The other intensional connective $\circ$ can be added by (R2)–(R3) as in the other logics. The modality of $\mathbf{E}$, especially rule ($\Box_4$), suggests that *all theorems* should be necessary. Hence, to add *truth*, $\boldsymbol{t}$, an extra axiom (A29) is needed beyond rules (R4) and (R5).

(A29) $(\boldsymbol{t} \to \boldsymbol{t}) \to \boldsymbol{t}$

The axiom can be viewed as stating that $\boldsymbol{t}$ is necessary (according to the definition of $\Box$). Alternatively, (A29) can be thought of as a wff obtained by *modus ponens* from an "exceptional instance" of restricted assertion, which is exceptional because $\boldsymbol{t}$ behaves as if it were an implicational wff. (A14)–(A16) suffice for negation without any interaction with $\Box$. All these extensions are *conservative*, though the resulting axiom set is known not to be independent (cf. Anderson *et al.* [1992, §R].)

E contains *classical logic* in the same sense as **T** does (cf. theorem 34). Rule (γ) is *admissible* in **E** (see theorem 35, Meyer and Dunn [1969]).

Some of the theorems that were mentioned in §3.9 equally apply to **E**. $\mathbf{E}_{\rightarrow}$ and $\mathbf{E}_{\overset{\sim}{\rightarrow}}$ are *decidable*, whereas $\mathbf{E}_{+}$ and **E** are *undecidable. Interpolation* fails both for $\mathbf{E}_{+}$ and **E** (and for $\mathbf{E}^{\circ\boldsymbol{t}}$). Nonetheless, antecedents and consequents of implicational theorems of **E** *share a propositional variable.*

An adequate (sound and complete) relational semantics for $\mathbf{E}^{\circ\boldsymbol{t}}$ can be defined by slightly amending a frame for $\mathbf{T}^{\circ\boldsymbol{t}}$. Let $\sqsubseteq$ be a partial order, and let the following new frame condition be included as (r24).

(r24) $\exists\iota \in I.\, R\alpha\iota\alpha.$

Then, $\mathbf{E}^{\circ\boldsymbol{t}}$ is sound and complete with respect to the resulting relational semantics. Mares [2000] uses semantical methods to show that the addition of classical negation to $\mathbf{E}^{\circ\boldsymbol{t}}$ is *not conservative.*

CE IS NOT CONSERVATIVE OVER **E** (MARES, 2000). *The formula* ($K_{\supset}$) *is a* theorem *of* **CE** *but* not *of* $\mathbf{E}^{\circ\boldsymbol{t}}$.

($K_{\supset}$) $\sim(((\sim p \vee q) \rightarrow (\sim p \vee q)) \rightarrow (\sim p \vee q)) \vee \sim((p \rightarrow p) \rightarrow p) \vee ((q \rightarrow q) \rightarrow q)$

Relevant implication: **R**

The first relevance logic we mentioned in this chapter was the "weak implicational calculus" of Church, that is, the implicational fragment of **R**. We end this section with briefly recalling some results concerning **R**.

The implicational fragment $\mathbf{R}_{\rightarrow}$ is axiomatized by (A1), (A2), (A4), (A18) and (R1). An equivalent *sequent calculus* is described in Dunn [1986, §3.6] together with the statement of the admissibility of cut and the proof of the *decidability* of $\mathbf{R}_{\rightarrow}$.

De Morgan negation is added by the axioms (A15)–(A16) — (A14) is provable. An adequate sequent calculus is in Dunn [1986, §3.7], and an *analytic tableaux* formulation is in McRobbie and Belnap [1979] (cf. also Anderson *et al.* [1992, §60]). $\mathbf{R}_{\overset{\sim}{\rightarrow}}$ is *decidable* — see Dunn [1986, §§3.6–3.7].

Positive **R** is axiomatized by the addition of the axioms (A7)–(A13) and rules (R2)–(R6) to $\mathbf{R}_{\rightarrow}$. An equivalent *Gentzen system* $\mathbf{LR}_{+}^{\circ\boldsymbol{t}}$ is in Dunn [1986, §3.10] and in Dunn [1973] (where it is denoted as $L\mathbf{R}^{+}$). A somewhat modified version is in Anderson and Belnap [1975, §28.5]. $\mathbf{R}_{+}^{\circ\boldsymbol{t}}$ is *undecidable* by Urquhart's result, that we mentioned in §3.9.

Urquhart demonstrates (s. Brady [2003, §11.6]) that the wff (with $p, q, r, s, t \in \mathbb{P}$)

$$\text{(fip)}\quad (q \circ (p \wedge (q \to r) \wedge (q \to (p \to q)) \wedge ((p \to q) \to q) \wedge ((q \wedge s) \to p))) \to \\ (((r \to (s \vee t)) \wedge (p \to p) \wedge (p \to t) \wedge (p \to (t \to t))) \to t)$$

together with the six-element De Morgan monoid, that was labeled CRYSTAL in Thistlewaite *et al.* [1988, 75] (see figure 2, p. 771), suffices to prove the *failure of interpolation* in positive **R**. The reasoning proceeds by defining an assignment to the variables so that the formula takes one of the two incomparable middle values. (Negation has fixed points on the two middle elements. The value of $\mathcal{A} \to \mathcal{B}$ is (i) $\boldsymbol{T}$ if $\mathcal{A} = \boldsymbol{F}$ or $\mathcal{B} = \boldsymbol{T}$, (ii) $\boldsymbol{F}$ if $\mathcal{B} < \mathcal{A}$ or $\mathcal{A} \parallel \mathcal{B}$, (iii) $\mathcal{B}$ if $\mathcal{A} = \boldsymbol{t}$, (iv) $\sim\mathcal{A}$ if $\mathcal{B} = \boldsymbol{f}$ and (v) $\mathcal{A}$ if $\mathcal{A} = \mathcal{B}$ and $\mathcal{B}$ is a or b.) Then, from the assumption that there is a suitable interpolant for (fip) it can be determined by direct calculations on the lattice that there is no suitable element in the monoid for the wff to take as a value.

An adequate set theoretical *semantics* for $\mathbf{R}_{+}^{\circ\boldsymbol{t}}$ was defined by Meyer and Routley [1972], and proofs of the soundness and completeness theorems can be also found in Dunn [1986, §4.7] as well as in Anderson *et al.* [1992, §48.3]. A frame can be defined as that for $\mathbf{T}^{\circ\boldsymbol{t}}$ — except, of course, the conditions which include * ((r7), (r8), (r9), (r17), and (r21)) — with the addition of the following condition.[44]

$$\text{(r25)}\quad R\alpha\beta\delta \wedge R\delta\gamma\eta \;\Rightarrow\; \exists\delta.\, R\alpha\gamma\delta \wedge R\delta\beta\eta.$$

Combining the negation-implication and the positive fragment gives **R** — *the logic of relevant implication.* A sound and complete *semantics* is defined by reversing the previous deletion, in other words, by (r1)–(r25). **R** has the *variable sharing* property which means that there are no implicational theorems in **R** without the antecedent and the consequent sharing some *information.*

$\mathbf{R}^{\circ\boldsymbol{t}}$ was *algebraized* by Dunn [1966], for expositions of this algebra see Anderson and Belnap [1975, §28.2] and Dunn [1986, §4.5]. The following is a definition of De Morgan monoids with a brief explanation.

DE MORGAN MONOIDS (DUNN, 1966). A *De Morgan monoid* $\mathfrak{D} = \langle D; \wedge, \vee, \circ, \sim, \boldsymbol{t}\rangle$ is an algebra (of type $\langle 2, 2, 2, 1, 0\rangle$) that is equationally characterized by (i)–(iii).

(i) $\langle D; \wedge, \vee, \sim\rangle$ is a distributive lattice with De Morgan negation;

(ii) $\langle D; \circ, \boldsymbol{t}\rangle$ is a square increasing Abelian monoid;

(iii) $\sim(a \circ \sim b) \circ a \leq b$ (where $\leq$ is the standard lattice order).

(i) means that $\langle D; \wedge, \vee, \sim\rangle$ is a *De Morgan lattice*, that is, the usual equations defining a distributive lattice are satisfied, and together with the unary operation

---

[44] In the presence of the other conditions the whole set of frame conditions may be simplified — (r25) is the condition that *corresponds* to axiom (A18). (The semantics obtained via modifying the one we gave above for $\mathbf{T}^{\circ\boldsymbol{t}}$ also differs in some other (small) details from Meyer and Routley's original semantics.)

$\sim$ the De Morgan laws hold. An Abelian monoid is an algebra with a binary associative operation, that is also commutative and has an identity element $\boldsymbol{t}$. Square increasingness is true when $\circ$ is "idempotent in one direction," so to speak, that is, $a \leq a \circ a$ is true. (iii) together with the De Morgan lattice provide residuation between $\circ$ and $\lambda xy.\ {\sim}(x \circ {\sim} y)$. This ensures further desirable interactions between the operations.

Translations of and into intuitionistic logic has been a favorite topic in logic ever since the invention of **H**. (We mentioned that a most prominent motivation behind linear logic is a translation of intuitionistic implication.) Though the question of the philosophical import of **H** (and of Heyting's arithmetic) for the foundations of mathematics has been outlined of course, early on by Gödel [1933a] and [1933b]. As Meyer [1973] shows *intuitionistic logic* can be translated into **R**. Let **H** be axiomatized with $\neg$ (intuitionistic negation) defined, rather that a primitive, operation. Any wff $\mathcal{B}$ contains finitely many propositional variables, let us say, $p_0, \ldots, p_n$. Then the translation of $\mathcal{B}$, $\tau(\mathcal{B})$, is inductively defined by (1)–(5).

(1) $\tau(\boldsymbol{F}) = \bigwedge_{i=0}^{n} p_i$;

(2) $\tau(p) = p$ where $p \in \mathbb{P}$;

(3) $\tau(\mathcal{A} \to \mathcal{B}) = \boldsymbol{t} \wedge \tau(\mathcal{A}) \to \tau(\mathcal{B})$;

(4) $\tau(\mathcal{A} \wedge \mathcal{B}) = \tau(\mathcal{A}) \wedge \tau(\mathcal{B})$;

(5) $\tau(\mathcal{A} \vee \mathcal{B}) = \tau(\mathcal{A}) \vee \tau(\mathcal{B})$.

Then, $\mathcal{B}$ is a theorem of **H** if and only if $\tau(\mathcal{B})$ is a theorem of $\mathbf{R}^{\boldsymbol{t}}$. Of course, since $\tau(\mathcal{B})$ is a positive wff, $\mathbf{R}^{\boldsymbol{t}}$ can be replaced by $\mathbf{R}^{\boldsymbol{t}}_{+}$. On the other hand, under the identity translation every theorem of $\mathbf{R}^{\boldsymbol{t}}_{+}$ is a theorem of **H**.

A potential next question is the relation of **R** to classical logic. We note before anything else that adding the principal type scheme of K to **R** collapses **R** into classical logic. Nowadays, it is a common place that classical logic has several functionally complete sets of connectives such as $\{\neg, \wedge\}$, $\{\,|\,\}$ and $\{\neg, \vee\}$, for instance. The wffs (1′)–(6′) (of section 1) all could be derived from $\mathcal{A} \vee \neg\mathcal{A}$ by a rule $\mathcal{A} \Rightarrow \mathcal{A} \vee \mathcal{B}$. This restates the observation that *classical theorems* can easily be translated into **R** theorems (cf. theorem 34). Of course, classically all theorems are equivalent, and so in a certain sense all classical theorems are equivalent to $\mathcal{A} \vee \neg\mathcal{A}$, which means that classical logic has $\aleph_0$ copies amongst the theorems of **R**. Moreover, $\aleph_0$ many instances of $\mathcal{A} \vee {\sim}\mathcal{A}$ are distinct in **R**, that is, they do not imply each other. Thus, classical logic is in **R** in another sense too. After all nothing is lost from classical logic in **R** in the sense of *theoremhood*; rather, because the implication of **R** is an *implication*, hence ${\sim}\mathcal{A} \wedge \mathcal{A} \to \mathcal{B}$, for instance, is *not* an **R** theorem (but ${\sim}\mathcal{A} \vee \mathcal{A} \vee \mathcal{B}$ is). An interesting turnaround is that classical first-order logic can be formalized with logical constants disjunction, negation and existential quantifier and with only *relevant inference* steps as the one sided Gentzen-style calculus K_1 in Dunn and Meyer [1989] demonstrates.

The relationship of **R** to classical logic has been investigated in a different, but also technically motivated setting. As we mentioned the Lindenbaum algebra of **R** is not a Boolean algebra, that is, **R** does not contain classical logic in the same way as well-known modal logics such as **K**, **B**, **S5**, etc. encompass classical logic. $\mathbf{R}_+$ can be extended to include $\neg$ (classical negation), without the resulting logic crumbling into the two-valued propositional system. The system of "classical relevance logic" **CR** is defined by the addition of axiom (A30) and rule (R15) (cf. Meyer and Routley [1973] and Meyer and Routley [1974]).

(A30) $\neg\neg\mathcal{A} \to \mathcal{A}$

(R15) $\mathcal{A} \wedge \mathcal{B} \to \neg\mathcal{C} \;\Rightarrow\; \mathcal{A} \wedge \mathcal{C} \to \neg\mathcal{B}$

Of course, $\neg$ is not $\sim$, thereby some of the undesirable (i.e., paradoxical) wffs piggyback on $\neg$ as theorems. For instance, $\mathcal{A} \wedge \neg\mathcal{A} \to \mathcal{B}$ as well as $\mathcal{A} \to \mathcal{B} \vee \neg\mathcal{B}$ are provable in **CR**. We reiterate that not only the relevance logics that formalize various reasonable properties of implication do not collapse into classical logic, but after adding classical negation to $\mathbf{R}_+$ the so obtained logic still stands apart from classical logic. Ultrafilters of a Boolean algebra form an antichain. Hence, the frame for **CR** which has a Boolean algebra (topped with further operations) as its Lindenbaum algebra can be simplified by turning $\sqsubseteq$ into identity. Meyer and Routley [1973] remark that the addition of Boolean negation might simplify the decidability problem for $\mathbf{R}_+$; and it seems it indeed did.

Another place where Boolean negation may be useful is in a *display logic* formulation of **R**. Display logic was introduced by Belnap [1982] (see also Anderson *et al.* [1992, §62]) as a generalization of Gentzen-style sequent calculi. Beyond the idea of having enough many structural connectives for different kinds of connectives (e.g., some extensional and intensional ones) that was already present in $L\mathbf{R}^+$, display logic builds in *residuation* into the syntax of a logic directly (on the level of the structural connectives). Display calculi enjoy both cut-free proofs and the subformula property as a result of exploiting the structural connectives for back and forth moves between the two sides of a consecution (which is what makes residuation attractive proof theoretically) without making the connectives (that are the counterparts of the structural connectives involved) part of a wff. **R** can be so Gentzenized, because the addition of classical negation does not yield new theorems in **R**'s language (see Anderson *et al.* [1992, §62.5.2]).

Some of the main points we made in this chapter are as follows. We argued that classical logic *does not* formalize implication or entailment, however, relevance logics *formalize* a whole variety of these notions. We presented some *technical details* that contribute not only to the proofs of certain lemmas and theorems, but are of interest in themselves and provide true philosophical insights.

We agree with Haack [1996] and Woods [2003], for instance, that certain formal results might have philosophical implications.[45] It seems that some philosophers are concerned about the place (or displacement) of classical logic in the late 20th and in the 21st century. In this chapter we tried to introduce a wide range of relevance logics so as to demonstrate their versatility. Some of the logics we mentioned, such as $\mathbf{T}_\to$, do not contain classical logic under any obvious translation, despite the fact that they *model* a certain kind of *entailment*. Some other logics, for example, $\mathbf{T}^{\circ\boldsymbol{t}}$, contain all the theorems of classical logic under a straightforward matching of the connectives without abandoning the requirement that *implications* should be *relevant*. Furthermore, some (though not each) of the relevance logics, but importantly **R**, can be extended to contain classical logic as an underlying Boolean algebra — without becoming a notational variant of classical logic.

---

[45] We find, however, Haack's classification of logics into "deviant/rivals" and "alternative/supplements" more that a bit "fuzzy."

It is our hope that the exploration of the relation of relevance logics to classical logic — that we slightly overemphasized in this chapter if looked at from purely the point of view of relevance logics — addresses some of the not fully uncommon philosophical concerns about the *advancement of nonclassical logics* including relevance logics. In a more technical tone, one has to profess that the Boolean extensions led to exciting logical results both for *intensional logics* in general, but also, specifically, for relevance logics without orthonegation.

Lastly, we want to point out that relevance logics connect fruitfully to further removed logics that are variously called *substructural*, *nonstandard*, *intensional* or *nonclassical*. However, the exploration of those relationships is way more that what could fit into the scope of this chapter.

ACKNOWLEDGEMENTS

I am grateful to Michael Dunn for advancing my knowledge of logic in a variety of ways for more that a decade, and for his useful comments on this chapter. I owe thanks to Alasdair Urquhart for his help at a crucial moment as the project of writing this chapter was getting on the road. I am thankful to the Robarts Library of the University of Toronto that generously gave me access to its collections (that led to some citations in this chapter). I wish to thank Steve Giambrone for acting as an intermediary in setting up the project. Financial support was provided by the School of Informatics at Indiana University in Bloomington, that is gratefully acknowledged.

BIBLIOGRAPHY

[Ackermann, 1956] W. Ackermann. Begründung einer strengen Implikation. *Journal of Symbolic Logic*, 21, 113–128, 1956.

[Allwein and Dunn, 1993] G. Allwein and J. M. Dunn. Kripke models for linear logic. *Journal of Symbolic Logic*, 58, 514–545, 1993.

[Anderson and Belnap, 1962] A. R. Anderson and N. D. Belnap. The pure calculus of entailment. *Journal of Symbolic Logic*, 27, 19–52, 1962.

[Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap. *Entailment. The Logic of Relevance and Necessity*, volume I. Princeton University Press, Princeton, NJ, 1975.

[Anderson *et al.*, 1992] A. R. Anderson, N. D. Belnap, and J. M. Dunn. *Entailment. The Logic of Relevance and Necessity*, volume II. Princeton University Press, Princeton, NJ, 1992.

[Anderson, 1960] A. R. Anderson. Entailment shorn of modality, (abstract). *Journal of Symbolic Logic*, 25, 388, 1960.

[Avron, 1988] A. Avron. The semantics and proof theory of linear logic. *Theoretical Computer Science*, 57, 161–184, 1988.

[Barwise, 1993] J. K. Barwise. Constraints, channels and the flow of information. In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, editors, *Situation Theory and its Applications*, pages 3–27. CSLI, Stanford, CA, 1993.

[Belnap, 1982] N. D. Belnap. Display logic. *Journal of Philosophical Logic*, 11, 375–417, 1982.

[Belnap, 1993] N. D. Belnap. Life in the undistributed middle. In K. Došen and P. Schroeder-Heister, editors, *Substructural Logics*, pages 31–41. Clarendon and Oxford University Press, Oxford, UK and New York, NY 1993.

[Bimbó and Dunn, 2002] K. Bimbó and J. M. Dunn. Four-valued logic. *Notre Dame Journal of Formal Logic*, 42, 171–192, 2002.

[Bimbó and Dunn, 2005] K. Bimbó and J. M. Dunn. Relational semantics for Kleene logic and action logic. *Notre Dame Journal of Formal Logic*, 46, 461–490, 2005.

[Bimbó, 2003] K. Bimbó. A set theoretical semantics for full propositional linear logic, (abstract). *Bulletin of Symbolic Logic*, 10, 136, 2003.

[Bimbó, 2005a] K. Bimbó. Types of I-free hereditary right maximal terms. *Journal of Philosophical Logic*, 34, 607–620, 2005.

[Bimbó, 2005b] K. Bimbó. Admissibility of cut in *LC* with fixed point combinator. *Studia Logica*, 81, 399–423, 2005.

[Boolos, 1995] G. Boolos. *The Logic of Provability.* Cambridge University Press, Cambridge, UK, 1995.

[Brady, 1984] R. T. Brady. Natural deduction systems for some quantified relevant logics. *Logique et Analyse*, 27, 355–377, 1984.

[Brady, 2003] R. T. Brady, editor. *Relevant Logics and their Rivals,* vol. II. *A continuation of the work of R. Sylvan, R. Meyer, V. Plumwood and R. Brady*, volume 59 of *Western Philosophy Series*. Ashgate, Burlington, VT, 2003.

[Church, 1951] A. Church. The weak theory of implication. In A. Menne, A. Wilhelmy, and H. Angsil, editors, *Kontrolliertes Denken, Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften*, pages 22–37. Komissions-Verlag Karl Alber, Munich, 1951.

[Cohen, 1963] P. Cohen. The independence of the continuum hypothesis, I. *Proceedings of the National Academy of Sciences*, 50, 1143–1148, 1963.

[Cohen, 1964] P. Cohen. The independence of the continuum hypothesis, II. *Proceedings of the National Academy of Sciences*, 51, 105–110, 1964.

[Curry and Feys, 1958] H. B. Curry and R. Feys. *Combinatory Logic*, volume I. North-Holland, Amsterdam, 1st edition, 1958.

[Curry, 1963] H. B. Curry. *Foundations of Mathematical Logic.* McGraw-Hill Book Company, New York, NY, 1963.

[Diamond, 1976] C. Diamond, editor. *Wittgenstein's Lectures on the Foundations of Mathematics.* Cornell University Press, Ithaca, NY, 1976.

[Dunn and Hardegree, 2001] J. M. Dunn and G. M. Hardegree. *Algebraic Methods in Philosophical Logic*, volume 41 of *Oxford Logic Guides.* Oxford University Press, Oxford, UK, 2001.

[Dunn and Meyer, 1989] J. M. Dunn and R. K. Meyer. Gentzen's cut and Ackermann's gamma. In J. Norman and R. Sylvan, editors, *Directions in Relevant Logic*, pages 229–240. Kluwer, Dordrecht, 1989.

[Dunn and Meyer, 1997] J. M. Dunn and R. K. Meyer. Combinators and structurally free logic. *Logic Journal of IGPL*, 5, 505–537, 1997.

[Dunn and Restall, 2002] J. M. Dunn and G. Restall. Relevance logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 6, pages 1–128. Kluwer, Amsterdam, 2nd edition, 2002.

[Dunn, 1966] J. M. Dunn. *The Algebra of Intensional Logics.* PhD thesis, University of Pittsburgh, Ann Arbor (UMI), 1966.

[Dunn, 1973] J. M. Dunn. A 'Gentzen system' for positive relevant implication, (abstract). *Journal of Symbolic Logic*, 38, 356–357, 1973.

[Dunn, 1976a] J. M. Dunn. Intuitive semantics for first-degree entailment and 'coupled trees'. *Philosophical Studies*, 29, 149–168, 1976.

[Dunn, 1976b] J. M. Dunn. A Kripke-style semantics for *R*-mingle using a binary accessibility relation. *Studia Logica*, 35, 163–172, 1976.

[Dunn, 1976c] J. M. Dunn. Quantification and *RM*. *Studia Logica*, 35, 315–322, 1976.

[Dunn, 1980] J. M. Dunn. A sieve for entailments. *Journal of Philosophical Logic*, 9, 41–57, 1980.

[Dunn, 1986] J. M. Dunn. Relevance logic and entailment. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 3, pages 117–229. Reidel, Dordrecht, 1st edition, 1986.

[Dunn, 1991] J. M. Dunn. Gaggle theory: An abstraction of Galois connections and residuation with applications to negation and various logical operations. In J. van Eijck, editor, *Logics in AI: European Workshop JELIA '90*, volume 478 of *Lecture Notes in Computer Science*, pages 31–51. Springer, Berlin, 1991.

[Dunn, 1993] J. M. Dunn. Partial-gaggles applied to logics with restricted structural rules. In K. Došen and P. Schroeder-Heister, editors, *Substructural Logics*, pages 63–108. Clarendon and Oxford University Press, Oxford, UK and New York, NY 1993.

[Dunn, 1995] J. M. Dunn. Gaggle theory applied to intuitionistic, modal and relevance logic. In I. Max and W. Stelzner, editors, *Logik und Mathematik. Frege-Kolloquium Jena*, pages 335–368. W. de Gruyter, Berlin, 1995.

[Dunn, 2001] J. M. Dunn. A representation of relation algebras using Routley-Meyer frames. In C. A. Anderson and M. Zelëny, editors, *Logic, Meaning and Computation. Essays in Memory of Alonzo Church*, volume 305 of *Synthese Library: Studies in Epistemology, Logic, Methodology and Philosophy of Science*, pages 77–108. Kluwer, Dordrecht, 2001.

[Enderton, 2001] H. B. Enderton. *A Mathematical Introduction to Logic.* Harcourt-Academic Press, Burlington, MA, 2nd edition, 2001.

[Fine, 1974] K. Fine. Models for entailment. *Journal of Philosophical Logic*, 3, 347–372, 1974.

[Frege, 1967] G. Frege. Begriffsschrift. In J. van Heijenoort, editor, *From Frege to Gödel*, pages 5–82. Harvard University Press, Cambridge, MA, 1967.

[Gentzen, 1935] G. Gentzen. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, 39, 176–210, 1935.

[Giambrone, 1985] S. Giambrone. TW_+ and RW_+ are decidable. *Journal of Philosophical Logic*, 14, 235–254, 1985.

[Girard, 1987] J.–Y. Girard. Linear logic. *Theoretical Computer Science*, 50, 1–102, 1987.

[Girard, 2004] J.–Y. Girard. Between logic and quantic: a tract. In T. Ehrhard, J.–Y. Girard, P. Ruet, and Ph. Scott, editors, *Linear Logic in Computer Science*, number 316 in London Mathematical Society Lecture Note Series, pages 346–381. Cambridge University Press, Cambridge, UK, 2004.

[Gödel, 1933a] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls, 1933. In S. Feferman, editor, *Collected Works*, volume I, pages 300–303. Oxford University Press and Clarendon Press, New York, NY and Oxford, UK, 1986.

[Gödel, 1933b] K. Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie, 1933. In S. Feferman, editor, *Collected Works*, volume I, pages 286–295. Oxford University Press and Clarendon Press, New York, NY and Oxford, UK, 1986.

[Gödel, 1938] K. Gödel. The consistency of the axiom of choice and of the generalized continuum hypothesis. *The Proceedings of the National Academy of Sciences*, 24, 665–557, 1938.

[Goldblatt, 2003] R. Goldblatt. Mathematical modal logic: a view of its evolution. *Journal of Applied Logic*, 1, 309–392, 2003.

[Haack, 1996] S. Haack. *Deviant Logic, Fuzzy Logic.* University of Chicago Press, Chicago, IL, 1996.

[Helman, 1977] G. H. Helman. *Restricted Lambda Abstraction and the Interpretation of some Nonclassical Logics.* PhD thesis, University of Pittsburgh, Ann Arbor (UMI), 1977.

[Hilbert and Ackermann, 1950] D. Hilbert and W. Ackermann. *Principles of Mathematical Logic.* Chelsea Publishing Company, New York, NY, 2nd edition, 1950.

[Hindley, 1997] J. R. Hindley. *Basic Simple Type Theory*, volume 42 of *Cambridge Tracts in Theoretical Computer Science.* Cambridge University Press, Cambridge, UK, 1997.

[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics.* Van Nostrand, New York, NY, 1952.

[Kopylov, 2001] A. P. Kopylov. Decidability of linear affine logic. In A. R. Meyer, editor, *Special issue: LICS 1995*, volume 164 of *Information and Computation*, pages 173–198. IEEE, 2001.

[Kripke, 1959] S. A. Kripke. The problem of entailment, (abstract). *Journal of Symbolic Logic*, 24, 324, 1959.

[Lemmon *et al.*, 1969] E. J. Lemmon, C. A. Meredith, D. Meredith, A. N. Prior, and I. Thomas. Calculi of pure strict implication. In J. W. Davis, D. J. Hockney, and W. K. Wilson, editors, *Philosophical Logic*, pages 215–250. D. Reidel, Dordrecht, 1969.

[Lincoln *et al.*, 1992] P. Lincoln, J. Mitchell, A. Scedrov, and N. Shankar. Decision problems for linear logic. *Annals of Pure and Applied Logic*, 56, 239–311, 1992.

[Mares and Meyer, 2001] E. D. Mares and R. K. Meyer. Relevant logics. In L. Goble, editor, *The Blackwell Guide to Philosophical Logic*, volume 4 of *Blackwell Philosophy Guides*, chapter 13, pages 280–308. Blackwell, Oxford, UK, 2001.

[Mares, 2000] E. D. Mares. **CE** is not a conservative extension of **E**. *Journal of Philosophical Logic*, 29, 263–275, 2000.

[Mares, 2002] E. D. Mares. Relevance logic. In D. Jacquette, editor, *A Companion to Philosophical Logic*, volume 22 of *Blackwell Companions to Philosophical Logic*, chapter 38, pages 609–627. Blackwell, Madden, MA, 2002.

[Mares, 2004] E. D. Mares. *Relevant Logic. A Philosophical Interpretation.* Cambridge University Press, Cambridge, UK, 2004.

[McRobbie and Belnap, 1979] M. A. McRobbie and N. D. Belnap. Relevant analytic tableaux. *Studia Logica*, 38, 187–200, 1979.

[Megill and Bunder, 1996] N. D. Megill and M. W. Bunder. Weaker **D**-complete logics. *Journal of IGPL*, 4, 215–225, 1996.

[Meredith and Prior, 1963] C. A. Meredith and A. N. Prior. Notes on the axiomatics of the propositional calculus. *Notre Dame Journal of Formal Logic*, 4, 171–187, 1963.

[Meyer and Dunn, 1969] R. K. Meyer and J. M. Dunn. E, R and γ. *Journal of Symbolic Logic*, 34, 460–474, 1969.

[Meyer and Routley, 1972] R. K. Meyer and R. Routley. Algebraic analysis of entailment I. *Logique et Analyse*, 15, 407–428, 1972.

[Meyer and Routley, 1973] R. K. Meyer and R. Routley. Classical relevant logics, I. *Studia Logica*, 32, 51–66, 1973.

[Meyer and Routley, 1974] R. K. Meyer and R. Routley. Classical relevant logics, II. *Studia Logica*, 33, 183–194, 1974.

[Meyer *et al.*, 1995] R. K. Meyer, M. A. McRobbie, and N. D. Belnap. Linear analytic tableaux. In P. Baumgartner, R. Hähnle, and J. Posegga, editors, *Theorem Proving with Analytic Tableaux and Related Methods*, volume 918 of *Lecture Notes in Computer Science*, pages 278–293, Berlin, 1995. 4th International Workshop, TABLEAUX'95, Springer.

[Meyer, 1966] R. K. Meyer. *Topics in Modal and Many-valued Logic.* PhD thesis, University of Pittsburgh, Ann Arbor (UMI), 1966.

[Meyer, 1973] R. K. Meyer. Intuitionism, entailment, negation. In H. Leblanc, editor, *Truth, syntax and modality*, pages 168–198. North-Holland, Amsterdam, 1973.

[Meyer, 2004] R. K. Meyer. Ternary relations and relevant logics. *Annals of Pure and Applied Logic*, 127, 195–217, 2004.

[Restall, 200+] G. Restall. Relevant and substructural logics. In D. M. Gabbay and J. Woods, editors, *The Handbook of the History of Logic*, (to appear), volume 6. Elsevier, Amsterdam, 200+.

[Routley and Meyer, 1972a] R. Routley and R. K. Meyer. The semantics of entailment, II. *Journal of Philosophical Logic*, 1, 53–73, 1972.

[Routley and Meyer, 1972b] R. Routley and R. K. Meyer. The semantics of entailment, III. *Journal of Philosophical Logic*, 1, 192–208, 1972.

[Routley and Meyer, 1973] R. Routley and R. K. Meyer. The semantics of entailment, I. In H. Leblanc, editor, *Truth, Syntax and Modality.* North-Holland, Amsterdam, 1973.

[Routley *et al.*, 1982] R. Routley, R. K. Meyer, V. Plumwood, and R. Brady. *Relevant Logics and their Rivals,* vol. I. Ridgeview Publishing Company, Atascadero, CA, 1982.

[Simpson, 1991] S. G. Simpson. *Subsystems of Second Order Arithmetic.* Springer, New York, NY, 1991.

[Smullyan, 1968] R. M. Smullyan. *First-order Logic.* Springer-Verlag, New York, NY, 1968, (Dover, New York, NY, 1995).

[Thistlewaite *et al.*, 1988] P. B. Thistlewaite, M. A. McRobbie, and R. K. Meyer. *Automated Theorem Proving in Non-classical Logics.* Pitman, London, UK, 1988.

[Urquhart, 1972] A. Urquhart. Semantics for relevant logic. *Journal of Symbolic Logic*, 37, 159–169, 1972.

[Urquhart, 1979] A. Urquhart. A topological representation theorem for lattices. *Algebra Universalis*, 8, 45–58, 1979.

[Urquhart, 1984] A. Urquhart. The undecidability of entailment and relevant implication. *Journal of Symbolic Logic*, 49, 1059–1073, 1984.

[Urquhart, 1990] A. Urquhart. The complexity of decision procedures in relevance logic I. In J. M. Dunn and A. Gupta, editors, *Truth or Consequences: Essays in honour of Nuel Belnap.* Kluwer, Amsterdam, 1990.

[Urquhart, 1993] A. Urquhart. Failure of interpolation in relevant logics. *Journal of Philosophical Logic*, 22, 449–479, 1993.

[Urquhart, 1999] A. Urquhart. The complexity of decision procedures in relevance logic II. *Journal of Symbolic Logic*, 64, 1774–1802, 1999.

[Urquhart, 2000] A. Urquhart. The complexity of linear logic with weakening. In S. R. Buss, P. Hájek, and P. Pudlák, editors, *Logic Colloquium'98*, volume 13 of *Lecture Notes in Logic*, pages 500–515, Natick, MA, 2000. A. K. Peters.

[Woods, 2003] J. Woods. *Paradox and Paraconsistency. Conflict Resolution in the Abstract Sciences.* Cambridge University Press, Cambridge, UK, 2003.

[Zermelo, 1962] E. Zermelo, editor. *Georg Cantor: Gesammelte Abhandlungen.* G. Olms Verlagsbuchhandlung, Hildesheim, 1962.

PARACONSISTENT LOGICS AND PARACONSISTENCY

Newton C. A. da Costa, Décio Krause and Otávio Bueno

1 INTRODUCTION

> "I predict a time when there will be mathematical investigations of calculi containing contradictions, and people will actually be proud of having emancipated themselves from contradictions."
>
> L. Wittgenstein

> "As a lightning clears the air of unpalatable vapors, so an incisive paradox frees the human intelligence from the lethargic influence of latent and unsuspected assumptions. Paradox is the slayer of Prejudice."
>
> J. J. Sylvester

In a few words, paraconsistent logics (PL) are the logics of inconsistent but nontrivial theories. A deductive theory is paraconsistent if its underlying logic is paraconsistent. A theory is inconsistent if there is a formula (a grammatically well-formed expression of its language) such that the formula and its negation are both theorems of the theory; otherwise, the theory is called consistent. A theory is trivial if all formulas of its language are theorems. Roughly speaking, in a trivial theory 'everything' (expressed in its language) can be proved. If the underlying logic of a theory is classical logic, or even any of the standard logical systems like intuitionistic logic, inconsistency entails triviality, and conversely. So, how can we speak of inconsistent but nontrivial theories? Of course, by changing the underlying logic to one which admits inconsistency without making the system trivial. Paraconsistent logics do just this job.

Our use of terms like 'consistency', 'inconsistency', 'contradictory' and similar ones is syntactical, which is in accordance with the original metamathematical terminology of Hilbert and his school. In order to treat such terms from a semantic point of view, in the field of paraconsistency, one must be able to build, first, a paraconsistent set theory. This is possible, as we will see, although most semantics for paraconsistent logics are classical, i.e., constructed inside classical set theories. So, to begin with, it is best to employ the above terms syntactically.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

1.1 The origins

The origins of paraconsistent logics go back to the first systematic studies dealing with the possibility of rejecting or restricting the law (or principle) of noncontradiction, which (in one of its possible formulations) says that a formula and its negation cannot both be true. The law of noncontradiction is one of the basic laws of traditional, or classical (Aristotelian), logic. This principle is important. After all, since inconsistency entails triviality, an inconsistent set of premises yields any well-formed statement as a consequence. The result is that the set of consequences of an inconsistent theory, or set of premises, will explode into triviality and the theory is rendered useless.

Another way of expressing this fact is by saying that under classical logic the closure of any inconsistent set of sentences includes every sentence. It is this which lies behind Popper's famous statement that the acceptance of inconsistency ''... would mean the complete breakdown of science" and that an inconsistent system is ultimately uninformative.[1]

Inconsistencies appear in various levels of discussion of science and philosophy. For instance, Peirce's world of 'signs' (which we inhabit) is an inconsistent and incomplete world. Bohr's theory of the atom is one of the well-known examples in science of an inconsistent theory. The old quantum theory of black-body radiation, Newtonian cosmology, the (early) theory of infinitesimals in the calculus, the Dirac δ-function, Stokes' analysis of pendulum motion, Michelson's 'single-ray' analysis of the Michelson-Morley interferometer arrangement, among others, can also be considered as cases of inconsistencies in science. Given cases such as these, it seems clear that we should not eliminate a priori inconsistent theories, but rather investigate them. In this context, paraconsistent logics acquire a fundamental role within science itself as well as in its philosophy. As we will see below, due to the wide range of applications which nowadays have been found for these logics, they have an important role in applied science as well.

The forerunners of paraconsistent logics are Jan Łukasiewicz and Nicolai I. Vasiliev. Independently of each other, both suggested in 1910 and 1911 that 'non-Aristotelian' logics could be obtained by rejecting the law of noncontradiction.[2] Although Łukasiewicz did not construct any system of paraconsistent logic, his ideas on the principle of noncontradiction in Aristotle influenced his student S. Jaśkowski in the construction of 'discussive' (or 'discursive') logic in 1948. (We will comment on Jaśkowski's systems below.) In 1911, 1912 and 1913, inspired by the works of Lobachewski on non-Euclidean geometry, initially called 'imaginary geometry', Vasiliev envisaged an *imaginary logic*, a non-Aristotelian logic where the principle of noncontradiction was not valid in general. According to Arruda, Vasiliev did not believe that there exist contradictions in the real world, but only in a possible world created by the human mind. Thus, he hypothesized imaginary worlds where the Aristotelian principles were not valid, even though Vasiliev did not develop his ideas in full.[3]

---

[1] For further details and references, see [da Costa and French, 2003], Chap. 5.

[2] For further historical details on PL, see [Arruda, 1980], [Arruda, 1989], [Arruda, 1990], [Bobenrieth, 1996], [da Costa and Marconi, 1989], [D'Ottaviano, 1990], [Puga and da Costa, 1988], [Guillaume, 1996], [da Costa *et al.*, 1995].

[3] Arruda systematized some points on Vasiliev's imaginary logic, giving rise to three systems of paraconsis-

The very first logician to construct a formal system of paraconsistent logic was Stanislaw Jaśkowski in 1948. His motivations came from his interests in systematizing theories that contain contradictions, such as dialectics, as well as to study theories where contradictions are caused by vagueness. He was also interested in the study of empirical theories whose postulates include contradictory assumptions (see section 4). Despite the wide range of possible applications, Jaśkowski's discussive logic was restricted to the propositional level. In 1958, the first author of this paper, independently of Jaśkowski, began the general study of contradictory systems [da Costa, 1958]. Ever since, da Costa has developed several systems related to paraconsistency (for instance, 'paraclassical logic' — see section 7.3), showing how to deal with inconsistencies from different perspectives. He apparently became the first logician to develop strong logical systems involving contradictions which could be useful for substantive parts of mathematics as well as the empirical and human sciences. It should be remarked that the adjective 'paraconsistent' (which means something like 'at the side of consistency') was suggested by F. Miró-Quesada, in 1976, in a letter to da Costa.

Already in the sixties, the interest in logics dealing with inconsistencies began in other parts of the world as well, particularly in Poland, Australia, United States, Italy, Argentina, Belgium, Ecuador, and Peru, mainly for its relations to da Costa's logics and to relevant and dialectical logic. Of course, in this paper, we cannot refer to all of these tendencies nor do justice to all the authors involved. For the historical details, we recommend the reading the papers mentioned above.

At least two facts have contributed to emphasize the relevance of these developments. The first is that in 1990, *Mathematical Reviews* added a new entry, 03B53, called 'Paraconsistent Logic'. From 2000 on, the title was changed to 'Logics admitting inconsistency (paraconsistent logics, discussive logics, etc.)', thus encompassing a wider subject. The second fact is that since 1996, several World Congresses on Paraconsistency have been organized.[4] Nowadays 'paraconsistency' can be regarded as a field of knowledge. But perhaps the most surprising fact concerning paraconsistent logic is related to its applications. As we will note later, there have been applications not only to the foundations of science and its philosophical analysis, but even to technology. Here we do not have space to present all the details, but the references list the original sources.

1.2 On the nature of logic

To understand the meaning and nature of logic, it is important to take into account that logic today is a field of knowledge of the same nature as mathematics. The results achieved in logic can be compared to those of mathematics and the empirical sciences in their depth and originality (let us just mention Gödel's theorems, the results in recursion theory and in the theory of models). And similarly to mathematics, we can divide logic into two domains: pure logic and applied logic. 'Pure' logic, similarly to pure mathematics, can be developed in principle *in abstracto*, independently of possible applications. In

---

tent logic [Arruda, 1977]. Nowadays, the logical works of Vasiliev have been studied in Russia, specially by V. A. Bazhanov.

[4] See `www.cle.unicamp.br/wcp3` for the page of the third congress.

particular, we can study paraconsistent logic or intuitionistic logic by themselves, basically with the aim of exploring their abstract mathematical properties. From this point of view, in developing a logical system, the logician can proceed as Hilbert suggested, when he said that "[t]he mathematician will have to take account not only of those theories that come near to reality, but also, as in geometry, of all logically possible theories" [Hilbert, 1902]. To sum up, from the pure viewpoint, logic studies certain abstract structures, such as formal languages, models, and Turing machines, independently of their applications.

Following Hilbert's suggestion, we can develop abstract (pure) systems where some principle of classical logic is violated, for instance, the principle that entails that from contradictory premises any formula can be derived; in symbols, $\alpha \wedge \neg\alpha \vdash \beta$. (The corresponding law, $(\alpha \wedge \neg\alpha) \rightarrow \beta$, is Duns Scotus Law, valid not only in classical logic, but in almost all known logical systems, including intuitionistic logic.)[5] This is the way taken by Vasiliev in the construction of his imaginary logics.

But to develop a logic we can also proceed from the *applied* point of view. In this case, we look at some domain of knowledge where it seems that some logic (in particular, a paraconsistent one) could be used to describe certain abstract structures that reflect the way certain deductive inferences are made in that domain. One of the best known examples is provided by Birkhoff and von Neumann's approach to quantum logic, when they insisted that quantum mechanics would demand a logic distinct from the classical one, giving rise to a whole new field of investigation, quantum logic (see [Dalla Chiara and Giuntini, 2001]).

It is sometimes claimed that nonclassical logics need to be developed *because* classical logic is wrong, and so it must be replaced by a suitable logic, in accordance to some philosophical criterion (for a discussion, see [Haack, 1974]). This would be the case, for example, of the intuitionistic Brouwer-Heyting logic if we consider it as a culmination of Brouwer's original philosophy of mathematics. Brouwer's stance implies that, in a certain sense, classical mathematics has basic shortcomings and that a constructive mathematics should take its place; the underlying logic of this constructive mathematics is a new one, different from classical logic. In particular, it may be argued that, in domains involving inconsistency (if really there are any), some other logic should be used instead of classical logic. Nowadays, there are also philosophers who believe that, in these fields, classical logic should be replaced by another logic (most of them think that the right logic would be relevant).

But this is not our view regarding the rejection of classical logic. We think that classical logic is a key subject that has, and will continue to have, strong interest and applications. The only difference is that, in certain domains, other logics, in particular paraconsistent logics, may be more adequate to make explicit some of the underlying structures that (apparently) are being assumed in these domains. Classical logic can't do that in every domain. This does not show that classical logic is wrong, but that its area of application should be restricted. The use of nonclassical logics in systematizing certain domains helps us understand important aspects of these domains. For example, with PL, the nature of negation has been better understood, and the significance of Russell set can be appreciated — see section 3.

---

[5] Some authors attribute this principle, also known as 'the principle of explosion', to Pseudo Scotus.

Furthermore, as will become clear below, in our view, classical logic is valid in its particular domain of application. With regard to PL, two possibilities emerge. In certain contexts, PL can be viewed as a 'heterodox' logic, as a 'rival' logic [Haack, 1974], that is, as a logic that deviates from classical logic with respect to some of its principles. But, in other contexts, PL can be viewed as a *supplement* to classical logic. After all, certain paraconsistent logics coincide with classical logic if we take into consideration just what are called 'well-behaved propositions' (roughly, those propositions that obey the principle of noncontradiction). In short, we don't intend to play according to PL rules alone. PL may be useful in some domains, as shown below, but we will continue to use classical logic, or any other logic for that matter, whenever it is convenient or appropriate.

So, we are inclined to agree with Gonseth [Gonseth, 1974, Chap. 8] and sustain that (applied) logic has an empirical counterpart. Nevertheless, we also have some reservations about Gonseth's picture. First, given our distinction between pure and applied logic, it is not necessary to eliminate a certain a priori aspect of logic, as Gonseth apparently wants (according to him, ''logic is the science of an arbitrary object"). This does not imply that we are endorsing the position that there is just one logic, which is independent of any domain of knowledge. As noted above, logic *can* be studied independently of any application, as a pure mathematical system, and in this sense, it can be considered as weakly a priori. So, even an applied logical system possesses an a priori dimension, in addition to its a posteriori one. In fact, we can begin by studying a logical system (say, some quantum logic) motivated by empirical considerations from science, but then proceed to verify whether this domain can be axiomatized, prove a completeness theorem for the resulting system, study other metalogical properties of the system, and so on.

From another point of view, however, logic deals with the underlying structures of inference of particular domains or theories. In this sense, a particular field (such as the quantum world, to continue with the example) may suggest that a different logic (that is, other than the classical one) is useful to accommodate certain features that cannot be dealt with by classical logic. For example, suppose we accept the view (advanced by E. Schrödinger, M. Born and others) that quantum objects are *non-individuals*, that is, they have no individuality since a quantum object is always indistinguishable of any other of a similar kind. It then seems that if we look at the quantum world as constituted by entities of this kind, classical logic (with Leibniz's Principle of the Identity of Indiscernibles) and classical mathematics (based on the notion of set, that is, collections of *distinguishable* objects) should be revised. This is particularly the case if we want to accommodate entities that, in certain contexts, can be regarded as 'individuals' (for instance, when an apparatus is prepared to work with particles), but in other contexts cannot be regarded as 'individuals' (say, when waves are taken into consideration). (Concerning these points, see [French and Krause, to appear].) So, different 'perspectives' of a domain of science may demand distinct logical tools, which puts us in a different framework than the classical one.

However, let us insist, the possibility of using nonstandard systems does not necessarily entail that classical logic is wrong, or that domains like quantum theory *need* at the moment another logic. Physicists and other scientists probably will continue to use classical (informal) logic in the near future. But we should realize that other forms of logic may help us understand certain features of these domains, which are not easily accommodated by classical means — as the concepts of non-individuality and complementarity in the quantum domain show (see [French and Krause, to appear], [da Costa and Krause, 2004], [da Costa and Krause, 2004a]).

We don't think there is just one 'true logic'. After all, distinct logical systems can be useful to describe different aspects of knowledge. (The same point can be made about distinct mathematical systems, and perhaps even about different physical physical systems.) In other words, we defend a form of *logical pluralism.* But our proposal is not relativist, since it's not the case that anything goes as far as applied logic is concerned. We can always rule out certain applied logical systems as being inadequate for certain domains. For instance, to capture constructive features of mathematical reasoning, classical logic is clearly inadequate; intuitionistic logic isn't. (We will return to the issue of logical pluralism below.) With regard to PL, we claim, with Granger [Granger, 1998], that paraconsistent logic can, and should, be employed in the development of certain domains, but only as a preliminary tool. In the end, classical logic may eventually replace it as the underlying logic of these domains. Our position does not exclude Granger's.

In summary, there are in principle various 'pure' logics whose potential applications depend not only on a priori reasons, but, above all, on the nature of the applications one has in mind. This is also true of PL.

This paper is organized as follows. In the next section, we present da Costa's C-logics. We then turn to paraconsistent set theories, and show, in particular, how they accommodate inconsistent objects, such as the Russell set. Next, we examine Jaśkowski's discussive logic, and show how it can be used in the formulation of the concept of partial truth. We then examine annotated logic, and some of its applications. Since all of these logics have been applied so extensively, such applications are here only touched upon (but references are given). Limitations of space also prevent us from examining all the related paraconsistent systems that have been developed in the literature. In fact, it would be impossible to do justice to all developments of PL and their corresponding applications. So, we limit ourselves to present the ideas and results with which our work is more closely related.

2 THE C-LOGICS

In this section, we study a class of logics termed C-logics, which show that it is possible to elaborate strong paraconsistent logical systems. In particular, within some of these systems, it is possible to build set theories and paraconsistent mathematics such that they contain standard mathematics. So, we can say that certain paraconsistent logico-mathematical systems increase the scope of traditional mathematics. In other words, it is possible to construct strong inconsistent systems without the immediate danger of trivialization. And there is no difficulty in reproducing within these systems the usual theories of logic and mathematics.

We begin with the propositional paraconsistent logic and, little by little, show how it is possible to get paraconsistent set theories and paraconsistent mathematics.

The way we approach the subject will be that of pure mathematics, in the same way that group theory or projective geometry of several dimensions are developed. The significance of all of this to the 'real world', that is, the possible applications of the resulting systems, will be discussed throughout the paper.

2.1 *The propositional calculi C_n*

Our initial goal is to develop propositional calculi that can be the basis of inconsistent but nontrivial theories.

We recall that a theory T, whose underlying logic is L and whose language is $\mathcal{L}$, is *inconsistent* if there is a formula α such that both α and $\neg\alpha$ (the negation of α) are theorems of T; otherwise, T is *consistent*. T is *trivial* if all formulas of $\mathcal{L}$ are theorems of T; otherwise, T is *nontrivial*. L is *paraconsistent* if it can be the underlying logic of inconsistent but nontrivial theories.

An expression of the form $\alpha \wedge \neg\alpha$, where $\wedge$ is the symbol for the conjunction, is called a *contradiction*. In general, if a theory is inconsistent, its logic enables us to derive, for whatever formula α, a contradiction $\alpha \wedge \neg\alpha$ from α and $\neg\alpha$; on the other hand, in most logics, from $\alpha \wedge \neg\alpha$ we can deduce both α and $\neg\alpha$. So, it is usual to call the trivial theories contradictory, which means that in such theories there are contradictions as theorems (or, equivalently, contradictory theorems).

We will begin by presenting the propositional calculus C_1. It seems natural that it should contain the usual connectives: $\rightarrow$ (implication), $\wedge$ (conjunction), $\vee$ (disjunction), and $\neg$ (negation); equivalence ($\leftrightarrow$) is defined as usual (see Definition 1 below). Furthermore, it seems also natural that C_1 should be composed of most all of the valid schemes and rules of classical propositional calculus, obeying the following conditions:[6]

I. In C_1 it should be not generally valid the principle of noncontradiction.

II. From two contradictory propositions, that is, one being the negation of the another, it should not be possible to deduce any proposition whatever.

The language $\mathcal{L}$ of C_1 contains the following primitive symbols: (i) propositional variables: a denumerable (infinite) set of propositional variables (formulas that are not analyzed at the propositional level); (ii) connectives: $\rightarrow$, $\wedge$, $\vee$ and $\neg$; and (iii) parentheses.

Formulas are defined as follows: (i) any propositional variable is a formula; (ii) if α and β are formulas, then $(\alpha \rightarrow \beta)$, $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$ and $\neg\alpha$ are formulas; (iii) the only formulas are those obtained from the preceding conditions (i) and (ii).

To facilitate the reading, we will adopt some conventions: (a) the symbol $\rightarrow$ is stronger than the others; (b) $\wedge$ and $\vee$ are stronger than $\neg$; and (c) external parentheses can be dispensed with.

DEFINITION 1. $\alpha \leftrightarrow \beta =_{\mathsf{def}} (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$

Now we list the postulates of C_1 (axiom schemes, axioms, and primitive inference rules). To begin with, we adopt the postulates of positive intuitionistic propositional logic:

---

[6] Until section 5, validity means syntactic validity: a formula is valid in a calculus if it has a proof in such calculus or if it is a theorem of this calculus.

1. $\alpha \to (\beta \to \alpha)$
2. $(\alpha \to \beta) \to ((\alpha \to (\beta \to \gamma)) \to (\alpha \to \gamma))$
3. $\alpha \wedge \beta \to \alpha$
4. $\alpha \wedge \beta \to \beta$
5. $\alpha \to (\beta \to \alpha \wedge \beta)$
6. $\alpha \to (\alpha \vee \beta)$
7. $\beta \to (\alpha \vee \beta)$
8. $(\alpha \to \gamma) \to ((\beta \to \gamma) \to (\alpha \vee \beta \to \gamma))$
9. $\alpha,\ \alpha \to \beta / \beta$

Let us now examine negation. We could think about adding to the postulates above the characteristic postulate of negation taken from the minimal calculus, namely,

$$(1) \quad (\alpha \to \beta) \to ((\alpha \to \neg\beta) \to \neg\alpha).$$

However, this would not be adequate. After all, from (1) and the postulates above, we can deduce the schema

$$\neg(\alpha \wedge \neg\alpha).$$

This stands for the principle of noncontradiction, which, by condition (I) above, should not be valid in C_1. Furthermore, in the minimal calculus, we can prove that the negation of any proposition can be derived from a contradiction, and this is not adequate either (see condition (II)). In the minimal calculus, we have (using (1))

$$\alpha, \neg\alpha, \beta \vdash \alpha \qquad \text{and} \qquad \alpha, \neg\alpha, \beta \vdash \neg\alpha;$$

hence

$$\alpha, \neg\alpha \vdash \neg\beta.$$

As a result, by applying twice the deduction theorem (which is a consequence of the postulates for implication of the minimal calculus), we have

$$\vdash \alpha \to (\neg\alpha \to \beta).$$

However, it seems interesting to accept (1), at least for those β such that

$$\neg(\beta \wedge \neg\beta).$$

This motivates us to adopt the following schema:

$$\neg(\beta \wedge \neg\beta) \to ((\alpha \to \beta) \to ((\alpha \to \neg\beta) \to \neg\alpha)).$$

We think it is convenient that every proposition be either *true* or *false*, at least in principle. (Nonetheless, we accept that there might be propositions that are true and whose negations are also true.) It then seems convenient to include, in addition, the following schema:

$$\alpha \vee \neg\alpha.$$

Now, $\neg(\alpha \wedge \neg\alpha)$ means that α satisfies the law of noncontradiction, that is, α is *well-behaved.* If this is not the case, that is, if $\alpha \wedge \neg\alpha$ holds, then α is *ill-behaved.* We then introduce the following definition:

DEFINITION 2. $\alpha^o =_{\mathsf{def}} \neg(\alpha \wedge \neg\alpha)$.

Let us now consider the schema

(2) $\neg\neg\alpha \rightarrow \alpha.$

We can reason heuristically as follows: if α is well-behaved, we can suppose that it obeys classical logic, and so (2) holds. If α is ill-behaved, then both α and $\neg\alpha$ are true, and by the postulates for implication, it follows that any proposition entails α; in particular, $\neg\neg\alpha$ entails α. Thus, (2) seems to be acceptable.

Finally, given what was said above, we also adopt postulates that entail that formulas built with well-behaved formulas are also well-behaved. That is:

$$\alpha^o \wedge \beta^o \rightarrow (\alpha \wedge \beta)^o, \quad \alpha^o \wedge \beta^o \rightarrow (\alpha \vee \beta)^o, \quad \alpha^o \wedge \beta^o \rightarrow (\alpha \rightarrow \beta)^o.$$

These new postulates can be written as follows:

$$\alpha^o \wedge \beta^o \rightarrow (\alpha \wedge \beta)^o \wedge (\alpha \vee \beta)^o \wedge (\alpha \rightarrow \beta)^o.$$

We can then deduce

$$\alpha^o \wedge \beta^o \rightarrow (\alpha \leftrightarrow \beta)^o.$$

Note that there is no necessity of assuming that

$$\alpha^o \rightarrow (\neg\alpha)^o,$$

for this can be proven from the postulates above.

So, we have the following list of postulates for C_1:

($\rightarrow_1$) $\alpha \rightarrow (\beta \rightarrow \alpha)$

($\rightarrow_2$) $(\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow (\alpha \rightarrow \gamma))$

($\rightarrow_3$) $\alpha,\ \alpha \rightarrow \beta / \beta$

($\wedge_1$) $\alpha \wedge \beta \rightarrow \alpha$

($\wedge_2$) $\alpha \wedge \beta \rightarrow \beta$

($\wedge_3$) $\alpha \rightarrow (\beta \rightarrow \alpha \wedge \beta)$

($\vee_1$) $\alpha \to (\alpha \vee \beta)$

($\vee_2$) $\beta \to (\alpha \vee \beta)$

($\vee_3$) $(\alpha \to \gamma) \to ((\beta \to \gamma) \to (\alpha \vee \beta \to \gamma))$

($\neg_1$) $\beta^o \to ((\alpha \to \beta) \to ((\alpha \to \neg\beta) \to \neg\alpha))$

($\neg_2$) $\alpha^o \wedge \beta^o \to (\alpha \wedge \beta)^o \wedge (\alpha \vee \beta)^o \wedge (\alpha \to \beta)^o$

($\neg_3$) $\alpha \vee \neg\alpha$

($\neg_4$) $\neg\neg\alpha \to \alpha$.

We will show shortly that C_1 has a bivalent semantics. But before doing that, let us study the main properties of C_1.

The concept of (formal) deduction of a formula from a set of formulas (i.e., in the standard notation, $\Gamma \vdash \alpha$) is defined in the usual way. In this case, we say that α is a syntactic consequence of the formulas in Γ. From now on, capital Greek letters stand for collections of formulas, while small Greek letters stand for formulas. We have, in C_1:

THEOREM 3.

(a) $\{\alpha\} \vdash \alpha$,

(b) $\Gamma \vdash \alpha$ entails $\Gamma \cup \Delta \vdash \alpha$,

(c) if $\Gamma \vdash \gamma$ for any $\gamma \in \Delta$ and $\Delta \vdash \alpha$, then $\Gamma \vdash \alpha$.

Proof. Immediate, from the standard definition of syntactic consequence ($\vdash$). ■

DEFINITION 4. α is a theorem of C_1 iff $\vdash \alpha$.

As usual, $\vdash \alpha$ means $\emptyset \vdash \alpha$. The symbols $\Rightarrow$ and $\Leftrightarrow$ are metalinguistic abbreviations of implication and bi-implication respectively.

THEOREM 5. *In* C_1*:*

(a) (Deduction theorem) $\Gamma \cup \{\alpha\} \vdash \beta \Rightarrow \Gamma \vdash \alpha \to \beta$

(b) (Modus ponens) $\{\alpha, \alpha \to \beta\} \vdash \beta$

(c) $\{\alpha,\beta\} \vdash \alpha \wedge \beta$, $\{\alpha,\beta\} \vdash \alpha$, $\{\alpha,\beta\} \vdash \beta$

(d) $\{\alpha\} \vdash \alpha \vee \beta$, $\{\beta\} \vdash \alpha \vee \beta$

(e) (Proof by cases) ($\Gamma \cup \{\alpha\} \vdash \gamma$ and $\Gamma \cup \{\beta\}) \vdash \gamma \Rightarrow \Gamma \cup \{\alpha \vee \beta\} \vdash \gamma$

(f) (Double negation elimination) $\{\neg\neg\alpha\} \vdash \alpha$.

Proof. As in classical logic. The deduction of the paraconsistent *reductio ad absurdum* is made using the rules for implication and the schema ($\neg_1$). ■

The next theorem is important, since it shows that conditions (I) and (II) that C_1 should obey are indeed satisfied.

THEOREM 6. *In C_1, the following schemas do not hold:*

1. $\neg\alpha \rightarrow (\alpha \rightarrow \beta)$
2. $\neg\alpha \rightarrow (\alpha \rightarrow \neg\beta)$
3. $\alpha \rightarrow (\neg\alpha \rightarrow \beta)$
4. $\alpha \rightarrow (\neg\alpha \rightarrow \neg\beta)$
5. $\alpha \wedge \neg\alpha \rightarrow \beta$
6. $\alpha \wedge \neg\alpha \rightarrow \neg\beta$
7. $(\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow \beta) \rightarrow \neg\alpha)$
9. $\alpha \rightarrow \neg\neg\alpha$
10. $(\alpha \leftrightarrow \neg\alpha) \rightarrow \beta$
11. $(\alpha \leftrightarrow \neg\alpha) \rightarrow \neg\beta$
12. $\neg(\alpha \wedge \neg\alpha)$
13. $\alpha \wedge \neg\alpha \rightarrow \neg(\alpha \wedge \neg\alpha)$

Proof. The result is established using the tables below, with 1 and 2 as designated values:

α	β	$\alpha \rightarrow \beta$	$\alpha \wedge \beta$	$\alpha \vee \beta$
1	1	1	1	1
2	1	1	1	1
3	1	1	3	1
1	2	1	1	1
2	2	1	1	1
3	2	1	3	1
1	3	3	3	1
2	3	3	3	1
3	3	1	3	3

α	$\neg\alpha$
1	3
2	1
3	1

■

We emphasize that, in C_1, the following principles — that play an important role in certain paradoxes, since they lead to trivialization — are *not* valid:

$$\neg\alpha \rightarrow (\alpha \rightarrow \beta),\ \neg\alpha \rightarrow (\alpha \rightarrow \neg\beta),\ \alpha \wedge \neg\alpha \rightarrow \beta,\ \alpha \wedge \neg\alpha \rightarrow \neg\beta.$$

THEOREM 7. *In C_1, the following version of reductio ad absurdum holds:*

$$(\Gamma \vdash \beta^o, \Gamma \cup \{\alpha\} \vdash \beta, \Gamma \cup \{\alpha\} \vdash \neg\beta) \Rightarrow \Gamma \vdash \neg\alpha.$$

Proof. By the deduction theorem and $(\neg_1)$. ■

COROLLARY 8. *In C_1, the following rules hold:*

$$(\Gamma \cup \{\alpha\} \vdash \beta^o, \Gamma \cup \{\alpha\} \vdash \beta, \Gamma \cup \{\alpha\} \vdash \neg\beta) \Rightarrow \Gamma \vdash \neg\alpha,$$

$$(\Gamma \cup \{\neg\alpha\} \vdash \beta^o, \Gamma \cup \{\neg\alpha\} \vdash \beta, \Gamma \cup \{\neg\alpha\} \vdash \neg\beta) \Rightarrow \Gamma \vdash \neg\alpha.$$

Proof. From the previous theorem and from the fact that in C_1 we have: $\vdash \alpha \vee \neg\alpha$ and $\vdash \neg\neg\alpha \rightarrow \alpha$. ■

THEOREM 9. *If we add to the axioms of C_1 the principle of noncontradiction, $\neg(\alpha \wedge \neg\alpha)$, as an additional postulate, we get the classical propositional calculus (CPC).*

Proof. By adding the principle of noncontradiction to C_1, we get $\vdash (\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow \neg\beta) \rightarrow \neg\alpha)$, and, hence, the postulate system for CPC presented by Kleene [Kleene, 1952, p. 82]. ■

From now on, to simplify the notation, we will eliminate the symbols { and } when we write deductions.

THEOREM 10. *In C_1, we have:*

1. $\alpha \rightarrow \alpha$
2. $\alpha \leftrightarrow \alpha$
3. $\alpha \rightarrow \beta, \beta \rightarrow \gamma \vdash \alpha \rightarrow \gamma$
4. $\alpha \rightarrow (\beta \rightarrow \gamma) \vdash \beta \rightarrow (\alpha \rightarrow \gamma)$
5. $\alpha \rightarrow (\beta \rightarrow \gamma) \vdash \alpha \wedge \beta \rightarrow \gamma$
6. $\alpha \wedge \beta \rightarrow \gamma \vdash \alpha \rightarrow (\beta \rightarrow \gamma)$
7. $\alpha \rightarrow \beta \vdash (\beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \gamma)$
8. $\alpha \rightarrow \beta \vdash (\gamma \rightarrow \alpha) \rightarrow (\gamma \rightarrow \beta)$
9. $\alpha \rightarrow \beta \vdash \alpha \wedge \gamma \rightarrow \beta \wedge \gamma$
10. $\alpha \rightarrow \beta \vdash \gamma \wedge \alpha \rightarrow \gamma \wedge \beta$
11. $\alpha \rightarrow \beta \vdash \alpha \vee \gamma \rightarrow \beta \vee \gamma$

12. $\alpha \to \beta \vdash \gamma \vee \alpha \to \gamma \vee \beta$

13. $\alpha \leftrightarrow \beta \vdash \beta \leftrightarrow \alpha$

14. $\alpha \leftrightarrow \beta, \beta \leftrightarrow \gamma, \gamma \vdash \alpha \leftrightarrow \gamma$

15. $\vdash (\alpha \leftrightarrow \beta) \leftrightarrow (\beta \leftrightarrow \alpha)$

16. $\vdash (\alpha \to (\alpha \to \beta)) \to (\alpha \to \beta)$

17. $\vdash (\alpha \to (\alpha \to (\alpha \to \beta))) \to (\alpha \to (\alpha \to \beta))$

18. $\vdash \underbrace{(\alpha \to \ldots (\alpha \to \beta) \ldots)}_{n+1\, times} \to \underbrace{(\alpha \to \ldots (\alpha \to \beta) \ldots)}_{n\, times}$

Proof. Immediate consequences of the way C_1 was constructed. ■

THEOREM 11. *Propositional positive intuitionistic logic is contained in* C_1.

Proof. This intuitionistic logic is, in fact, characterized by the postulates $(\neg_1)$ to $(\vee_3)$. ■

THEOREM 12. *In* C_1, *we have:*

1. $\beta^o, \alpha \to \beta \vdash \neg\beta \to \neg\alpha$

2. $\beta^o, \alpha \to \neg\beta \vdash \beta \to \neg\alpha$

3. $\beta^o, \neg\alpha \to \beta \vdash \neg\beta \to \alpha$

4. $\beta^o, \neg\alpha \to \neg\beta \vdash \beta \to \alpha$

5. $(\alpha \to \neg\alpha) \to \neg\alpha$

6. $(\neg\alpha \to \alpha) \to \alpha$

Proof. We prove item 3. We have: $\beta^o, \neg\alpha \to \beta, \neg\alpha \vdash \beta^o$; $\beta^o, \neg\alpha \to \beta, \neg\beta, \neg\alpha \vdash \neg\beta$; $\beta^o, \neg\alpha \to \beta, \neg\beta, \neg\alpha \vdash \beta$. Hence, $\beta^o, \neg\alpha \to \beta, \neg\beta \vdash \neg\neg\alpha$, therefore $\beta^o, \neg\alpha \to \beta, \neg\beta \vdash \alpha$. So, $\beta^o, \neg\alpha \to \beta \vdash \neg\beta \to \alpha$. ■

Remark: Since C_1 has several of the usual properties of classical propositional calculus, we can prove various other schemas as well, as is easy to see.

THEOREM 13. C_1 *is a sub-calculus of the classical propositional calculus.*

Proof. All postulates of C_1 are valid in the classical propositional calculus. Furthermore, Theorem 6 shows that there are schemas that are valid in the latter calculus, but which are not valid in C_1. ■

THEOREM 14. *In C_1, the following schemas are not valid:*

$$(\alpha \wedge \beta) \wedge \neg\alpha \rightarrow \beta \ \ \textit{and} \ \ \alpha \vee \beta \rightarrow (\neg\alpha \rightarrow \beta).$$

Proof. If these schemas were valid in C_1, then using the deduction theorem, modus ponens and the schema $\alpha \rightarrow \alpha \vee \beta$, we could derive in C_1 the schema $\alpha \rightarrow (\neg\alpha \rightarrow \beta)$. ■

The previous theorem could also be proven using the tables above. As a result, several other schemas that are valid in classical logic are not theorems of C_1.

COROLLARY 15. *The following rules are not valid in C_1 (disjunctive syllogisms):*

$$\frac{\alpha \vee \beta, \neg\alpha}{\beta} \ \ \textit{and} \ \ \frac{\neg\alpha \vee \beta, \alpha}{\beta}.$$

THEOREM 16. *If $\alpha_1, \ldots, \alpha_n$ are the prime components (propositional variables) of a formula α, then a necessary and sufficient condition for α to be provable in the classical propositional calculus is that $\alpha_1^o, \ldots, \alpha_n^o \vdash \alpha$ in C_1.*

Proof. If $\alpha_1^o, \ldots, \alpha_n^o \vdash \alpha$ in C_1, then $\vdash \alpha$ in the classical propositional calculus, since β^o is an abbreviation of $\neg(\beta \wedge \neg\beta)$. Now, if $\vdash \alpha$ in the classical propositional calculus (supposed axiomatized as in [Kleene, 1952]), then there exists a proof P of α in this calculus, in which only those formulas whose prime components are among $\alpha_1, \ldots, \alpha_n$ appear. So, if k is one of the formulas in P, given postulate $(\neg_2)$, we have that $\alpha_1^o, \ldots, \alpha_n^o \vdash k^o$ in C_1. Furthermore, in C_1, we have that $\alpha_1^o, \ldots, \alpha_n^o \vdash (\gamma \rightarrow \delta) \rightarrow ((\gamma \rightarrow \neg\delta) \rightarrow \neg\gamma)$, with the usual restrictions. But, since every postulate of the classical propositional calculus of [Kleene, 1952] is valid in C_1, with the exception of $(\theta \rightarrow \pi) \rightarrow ((\theta \rightarrow \neg\pi) \rightarrow \neg\theta)$, we see that P can be transformed into a deduction, in C_1, of α from $\alpha_1^o, \ldots, \alpha_n^o$. ■

THEOREM 17. *If $\alpha_1, \ldots, \alpha_n$ are the prime components (propositional variables) of the formulas Γ and of a formula α, then a necessary and sufficient condition for $\Gamma \vdash \alpha$ in the classical propositional calculus is that $\Gamma, \alpha_1^o, \ldots, \alpha_n^o \vdash \alpha$ in C_1.*

Proof. Analogous to the previous one. ■

THEOREM 18 (A. I. Arruda). *In C_1, we have: $\vdash \alpha^{oo}$.*

Proof. $\vdash \alpha^{oo}$ means $\neg(\alpha^o \wedge \neg\alpha^o)$, that is, $\neg(\alpha^o \wedge \neg\neg(\alpha \wedge \neg\alpha))$. But $\alpha^o \wedge \neg\neg(\alpha \wedge \neg\alpha) \vdash \alpha^o$ and $\alpha^o \wedge \neg\neg(\alpha \wedge \neg\alpha) \vdash \alpha \wedge \neg\alpha$, hence $\vdash \neg(\alpha^o \wedge \neg\neg(\alpha \wedge \neg\alpha))$, that is, $\vdash \alpha^{oo}$. ■

THEOREM 19. *In C_1, we have: $\vdash \alpha^o \rightarrow (\neg\alpha)^o$.*

Proof. We have: $\alpha^o, \neg\alpha \wedge \neg\neg\alpha \vdash \alpha$ and $\alpha^o, \neg\alpha \wedge \neg\neg\alpha \vdash \neg\alpha$. Given that $\alpha^o, \neg\alpha \wedge \neg\neg\alpha \vdash \alpha^o$, it follows that $\alpha^o \vdash \neg(\neg\alpha \wedge \neg\neg\alpha)$, and hence, $\alpha^o \vdash (\neg\alpha)^o$. Thus, $\vdash \alpha^o \rightarrow (\neg\alpha)^o$. ■

DEFINITION 20 (Strong negation). $\neg^\star\alpha =_{\mathsf{def}} \neg\alpha \wedge \alpha^o$

THEOREM 21. *In C_1, $\vdash (\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow \neg^\star\beta) \rightarrow \neg^\star\alpha)$*

Proof. We have

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \alpha^o \vdash \neg^\star\alpha, \text{ and} \tag{1}$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \neg\alpha^o \vdash \alpha \wedge \neg\alpha. \text{ But}$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \alpha \wedge \neg\alpha \vdash \beta,$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \alpha \wedge \neg\alpha \vdash \neg\beta \text{ and}$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta \vdash \alpha^o. \text{ Therefore,}$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \neg\alpha^o \vdash \alpha,$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \neg\alpha^o \vdash \neg\alpha,$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \neg\alpha^o \vdash \alpha^o, \text{ and}$$

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta,\, \neg\alpha^o \vdash \neg^\star\alpha. \tag{2}$$

From (1) and (2), using proof by cases and excluded middle, we have that

$$\alpha \to \beta,\, \alpha \to \neg^\star\beta \vdash \neg^\star\alpha, \text{ hence}$$

$$\vdash (\alpha \to \beta) \to ((\alpha \to \neg^\star\beta) \to \neg^\star\alpha).$$

■

THEOREM 22. $\vdash \alpha \to (\neg^\star\alpha \to \beta)$

Proof. We have:

$$\alpha,\, \neg\alpha \wedge \alpha^o,\, \neg\beta \vdash \alpha,$$

$$\alpha,\, \neg\alpha \wedge \alpha^o,\, \neg\beta \vdash \neg\alpha,$$

$$\alpha,\, \neg\alpha \wedge \alpha^o,\, \neg\beta \vdash \alpha^o. \text{ Hence,}$$

$$\alpha,\, \neg\alpha \wedge \alpha^o \vdash \neg\neg\beta,$$

$$\alpha,\, \neg\alpha \wedge \alpha^o \vdash \beta \text{ and } \alpha,\, \neg^\star\alpha \vdash \beta. \text{ So,}$$

$$\vdash \alpha \to (\neg^\star\alpha \to \beta).$$

■

THEOREM 23. $\vdash \alpha \vee \neg^\star\alpha$

Proof. We have:

$$\vdash (\alpha \vee \neg^\star\alpha) \leftrightarrow \alpha \vee (\neg\alpha \wedge \alpha^o),$$

$$\vdash (\alpha \vee \neg^\star\alpha) \leftrightarrow (\alpha \vee \neg\alpha) \wedge (\alpha \vee \alpha^o), \text{ and}$$

$$\vdash (\alpha \vee \neg^{\star}\alpha) \leftrightarrow (\alpha \vee \alpha^{o}).$$

But i) $\alpha^{o} \vdash \alpha \vee \alpha^{o}$ and ii) $\neg\alpha^{o} \vdash \alpha \wedge \neg\alpha \vdash \alpha \vdash \alpha \vee \alpha^{o}$. So, from i) and ii) it follows that $\vdash \alpha \vee \alpha^{o}$, hence $\vdash \alpha \vee \neg^{\star}\alpha$. ■

THEOREM 24. *The connectives* $\rightarrow$, $\wedge$, $\vee$ *and* $\neg^{\star}$, *in* C_1, *satisfy all schemas and inference rules of classical propositional calculus.*

Proof. In C_1, the following schemas are valid:

$$(\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow \neg^{\star}\beta) \rightarrow \neg^{\star}\alpha),$$

$$\alpha \rightarrow (\neg^{\star}\alpha \rightarrow \beta) \text{ and}$$

$$\alpha \vee \neg^{\star}\alpha.$$

If we add to these schemas the postulates ($\rightarrow_1$) to ($\vee_3$) above, we get the postulates for the classical propositional calculus (see [Kleene, 1952]). ■

THEOREM 25. *In* C_1, *Peirce's Law is valid; that is,*

$$\vdash ((\alpha \rightarrow \beta) \rightarrow \alpha) \rightarrow \alpha.$$

Proof. Peirce's Law can be derived in the classical propositional calculus. So, by the previous theorem, it is provable in C_1. ■

THEOREM 26. *In* C_1, *we have:* $\vdash \alpha \vee (\alpha \rightarrow \beta)$.

Proof. $\vdash \alpha \vee (\alpha \rightarrow \beta)$ is valid in the classical propositional calculus. ■

The results above show that classical propositional calculus is, in a certain sense, contained in C_1, although C_1 is a sub-calculus of it.

DEFINITION 27. A schema or formula α *trivializes* a calculus C if by adding α to C, the new calculus is trivial, that is, all of its formulas are theorems.

THEOREM 28. *Any formula of the form* $\alpha \wedge \neg^{\star}\alpha$ *trivializes* C_1.

Proof. By the results above, we see that $\neg^{\star}$ is a classical negation. So, $\alpha \wedge \neg^{\star}\alpha$ entails any formula, that is, $\vdash \alpha \wedge \neg^{\star}\alpha \rightarrow \beta$. ■

THEOREM 29. *The schema* $\alpha \leftrightarrow (\alpha \rightarrow \beta)$, *where* β *is any formula, trivializes* C_1.

Proof. In C_1, we have (i) $\vdash \alpha \vee (\alpha \rightarrow \beta)$. Hence, from α and $\alpha \leftrightarrow (\alpha \rightarrow \beta)$, we deduce β. From $\alpha \rightarrow \beta$ and from $\alpha \leftrightarrow (\alpha \rightarrow \beta)$, we deduce β again. So, from (i), $\vdash \beta$. ■

Since, in C_1, we have that: $\vdash \alpha \vee \underbrace{(\alpha \to (\alpha \to \ldots (\alpha \to \beta) \ldots)}_{n \text{ occurences of } \alpha}$, then the schema

$$\alpha \leftrightarrow \alpha \vee (\alpha \to (\alpha \to \ldots (\alpha \to \beta) \ldots),$$

where β is any formula, also trivializes C_1.

THEOREM 30. *In* C_1,

$$\nvdash (\alpha \leftrightarrow (\alpha \vee \alpha)) \leftrightarrow (\neg\alpha \leftrightarrow (\neg(\alpha \vee \alpha)).$$

Proof. Use the table given at Theorem 2.1.3, and assign to α the value 2. ■

THEOREM 31. *In* C_1*, the following schema and rule are not valid:*

$$(\alpha \leftrightarrow \beta) \to ((\neg\alpha \leftrightarrow \neg\beta), \quad \frac{\alpha \leftrightarrow \beta}{\neg\alpha \leftrightarrow \neg\beta}.$$

Proof. Immediate consequence of the previous theorem. ■

2.2 *The hierarchy* C_n, $0 \leq n \leq \omega$

The calculus C_1 is not the only that satisfies the conditions I and II formulated above (see page 797). Among other possible solutions, we will indicate, in what follows, a hierarchy of calculi which satisfy such conditions, except for the first one, which will be taken to be the classical propositional calculus. The hierarchy is the following:

(3) $C_0, C_1, C_2, \ldots, C_n, \ldots, C_\omega$,

where C_0 is the classical propositional calculus and the remaining ones are defined below.

To begin with, let us introduce the following definition:

DEFINITION 32.

(i) $\alpha^{(1)}$ stands for α^o

(ii) $\alpha^{(n)}$ stands for $\alpha^{n-1} \wedge (\alpha^{(n-1)})^o$, $2 \leq n \leq \omega$.

The calculus C_n ($0 < n < \omega$) is then individualized by the postulates $(\to_1)$ to $(\vee_3)$ above, plus $\alpha \vee \neg\alpha$, $\neg\neg\alpha \to \alpha$ and the following ones:

(n_1) $\beta^{(n)} \to ((\alpha \to \beta) \to ((\alpha \to \neg\beta) \to \neg\alpha)$

(n_2) $\alpha^{(n)} \wedge \beta^{(n)} \to ((\alpha \to \beta)^{(n)} \wedge (\alpha \wedge \beta)^{(n)} \wedge (\alpha \vee \beta)^{(n)})$.

The calculus C_ω has as postulates $(\rightarrow_1)$ to $(\vee_3)$ plus $\alpha \vee \neg\alpha$ and $\neg\neg\alpha \rightarrow \alpha$.

We can see that, in $C_n, 1 \leq n \leq \omega$, we can substitute the schema $\beta^{(n)} \wedge \beta \wedge \neg\beta \rightarrow \gamma$ for the postulate (n_1). It is also easy to verify that the results above, proven for C_1, with suitable adaptations, can be also proven for $C_n, n = 2, 3, \ldots$

THEOREM 33. *Each calculi in the hierarchy (3) is strictly stronger than those which follow it.*

Proof. (A. I. Arruda) The tables of the proof of Theorem 6, which we call here T_1, show that C_0 properly contains C_1. With the aim of proving that C_i properly contains C_{i+1}, $i = 1, 2, \ldots$, we define the following tables $T_2, T_3, \ldots$ (note that they do not include the tables for negation, which will be defined separately). T_2 is obtained from T_1 (as in the proof of Theorem 6) by the addition of a new value, 4, which will be the only non-designated value; we obtain T_n from T_{n-1} by adding a new value $n + 2$, which will be the only non-designated one etc. The tables T_n are then obtained from T_{n-1}, $n = 2, 3, \ldots$, as follows. (Note that the rules below refer to the new arrangements of values, which result from the addition of a new value, and no changes are made, in the new tables, of those already obtained in the tables of order $n - 1$.)

(1) *Conjunction*: If the components have different values, the conjunction will have the greatest value among the values of its conjuncts; if the values are equal, this will be the value of the conjunction.

(2) *Disjunction*: If the values of the components are distinct, the disjunction will have the smaller value among the values of the components; if they are equal, this value will be the value of the disjunction.

(3) *Implication*: If the values of the components are distinct, the implication will have the value of the consequent; if they are equal, the implication will have the value 1.

As for negation, T_n gives us the following table:

α	1	2	3	$\cdots$	n	n+1	n+2
$\neg\alpha$	n+2	1	2	$\cdots$	n-1	n	1

In the table above, $n = 1, 2, \ldots, n, n + 1$ are the designated values and $n + 2$ is the only non-designated one.

The tables T_n show that C_{n-1} strictly contains C_n. For instance, the postulate $\beta^{(n-1)} \rightarrow ((\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow \neg\beta) \rightarrow \neg\alpha))$ and the schema $\alpha^{(n-1)} \rightarrow (\neg\alpha)^{(n-1)}$ are not valid in C_n, which may be verified with some work ($n > 1$). Furthermore, C_ω is strictly weaker than all the other calculi of the hierarchy. ■

2.3 Theories

In this section, we will make explicit reference to the calculus C_1, although the exposition can be extended to all the calculi $C_n, 0 \leq n \leq \omega$.

DEFINITION 34. $\overline{\Delta} =_{\text{def}} \{\alpha : \Delta \vdash \alpha\}$

THEOREM 35.

(i) $\Delta \subset \overline{\Delta}$

(ii) $\Delta \subset \Gamma \Rightarrow \overline{\Delta} \subset \overline{\Gamma}$

(iii) $\overline{\overline{\Delta}} \subset \overline{\Delta}$

Proof. Immediate. ■

DEFINITION 36. Δ is a theory if and only if $\overline{\Delta} = \Delta$.

DEFINITION 37. Let Υ be the set of all formulas of C_1. Then Δ is *inconsistent* if there exists α such that $\Delta \vdash \alpha$ and $\Delta \vdash \neg\alpha$, otherwise Δ is *consistent.* Γ is *trivial* if $\overline{\Gamma} = \Upsilon$; otherwise, Γ is nontrivial. An inconsistent set is also called *contradictory.* A theory Δ is *paraconsistent* if it is inconsistent but nontrivial.

DEFINITION 38. Expressions of either the form $\alpha \wedge \neg\alpha$ or $\neg\alpha \wedge \alpha$ are called *contradictions.*

It is easy to see that Δ is inconsistent if and only if $\Delta \vdash \alpha \wedge \neg\alpha$ for some formula α.

THEOREM 39. *C_1 can be the underlying calculus of paraconsistent theories.*

Proof. It suffices to consider the theory $\overline{\{\alpha \wedge \neg\alpha\}}$, where α is a propositional variable, and to apply the tables of Theorem 6. ■

DEFINITION 40. A formula α trivializes a calculus C if, by adding α to C as a new axiom, the new resulting system is trivial. In this case, C is said to be *finitely trivializable.*

For instance, the intuitionistic or classical implicative propositional calculi and the classical positive propositional calculus are not finitely trivializable, while the classical predicate calculus is.

THEOREM 41. *A formula of the form $\alpha \wedge \neg\alpha \wedge \alpha^{(n)}$ trivializes C_n, $1 \leq n < \omega$.*

Proof. Immediate. ■

THEOREM 42. *C_ω is not finitely trivializable.*

Proof. Let us consider the following matrix M, where 1 is designated:

α	β	$\alpha \rightarrow \beta$	$\alpha \wedge \beta$	$\alpha \vee \beta$
0	0	1	0	0
1	0	0	0	1
0	1	1	0	1
1	1	1	1	1

If α does not begin with the symbol $\neg$, then:

1) If the value of α is 1, then the value of $\neg^n\alpha$ is 1, where $\neg^n$ stands for $\neg\neg\ldots\neg$, ($\neg$ repeated n times, $n \geq 1$).

2) If the value of α is 0, then the value of $\neg^{2k}\alpha$ is 0, and the value of $\neg^{2k+1}\alpha$ is 1, for all $k = 0, 1, 2, \ldots$

M is sound for C_ω, as is easy to see. By induction on the length of the formulas, we can show that no formula assumes the value 0. So, there is no formula γ such that $\gamma \vdash \alpha$ (or $\vdash \gamma \to \alpha$), for any formula α of C_ω. ■

The following results can also be proven without difficulty:

THEOREM 43. *The calculi C_n, $0 < n < \omega$, with a finite number of propositional variables, are trivializable by formulas of the form $\alpha \wedge \neg\alpha \wedge \alpha^{(n)}$.*

THEOREM 44. *C_ω, with a finite number of propositional variables, cannot be finitely trivializable.*

It should be remarked that if a theory is based on C_n, there is more risk of trivialization than if it were based on C_{n+1}. To obtain the 'maximum security' in avoiding trivialization, we should adopt C_ω. But the further we go in the hierarchy (3), the weaker the calculi we get.

THEOREM 45 (A. I. Arruda). *In C_n, $1 \leq n \leq \omega$, there is no reduction of negations; that is, expressions like $\neg^p\alpha \leftrightarrow \neg^q\alpha$, for $p \neq q$ ($p, q = 0, 1, 2, \ldots$), are not valid in these calculi.*

Proof. It suffices to use the following truth-tables: the values of the tables are the integers ≥ 1 and the only non-designated value is 1. If d stands for designated value and $v(\alpha)$ is the value of α, then the tables for $\to$, $\wedge$, $\vee$ and $\neg$ are as follows:

$\to$: $v(\alpha \to \beta) = 1$ iff $v(\alpha) = d$ and $v(\beta) = 1$; $v(\alpha \to \beta) = 2$ otherwise.

$\wedge$: $v(\alpha \wedge \beta) = 2$ iff $v(\alpha) = d$ and $v(\beta) = d$, and $v(\alpha \wedge \beta) = 1$ otherwise.

$\vee$: $v(\alpha \vee \beta) = 1$ iff $v(\alpha) = 1$ and $v(\beta) = 1$, and $v(\alpha \vee \beta) = 2$ otherwise.

$\neg$: $v(\neg\alpha) = 2$ iff $v(\alpha) = 1$ and $v(\neg\alpha) = 1$ otherwise.

■

COROLLARY 46 (A. I. Arruda). *The calculi C_n, $1 \leq n \leq \omega$, do not have finite characteristic truth-tables; that is, finite tables such that a necessary and sufficient condition for α to be a theorem is that it assumes only designated values in these tables.*

Proof. From the hypothesis of the corollary, it is easy to see that if the calculi had characteristic finite tables, they would enable reduction of negations. ■

2.4 Quantification

Corresponding to the hierarchy (3), we construct the corresponding first-order predicate calculi. These new calculi will be denoted by

(4) $C_0^\star, C_1^\star, C_2^\star, \ldots, C_n^\star, \ldots, C_\omega^\star$.

To begin with, we will construct the calculus $C_1^\star$. Let $\mathcal{L}^\star$ be a first-order language that contains the usual symbols, connectives and quantifiers, as well as a denumerable family of individual variables, predicate symbols (one for each arity), and auxiliary symbols. We could also suppose that $\mathcal{L}^\star$ contains functional symbols, and in particular, individual constants; but we will not make this supposition here, except when explicitly mentioned. The notions of formula, free and bound variables in a formula, sentence (formula without free variables) etc. are standard. The notations and metalogical conventions extend those made for the propositional calculi.

$C_0^\star$ is the classical first-order predicate calculus. The postulates of $C_1^\star$ are those of C_1 (suitably adapted) plus the following:

(I) $\frac{\alpha \to \beta(x)}{\alpha \to \forall x \beta(x)}$

(II) $\forall x \alpha(x) \to \alpha(y)$

(III) $\alpha(x) \to \exists x \alpha(x)$

(IV) $\frac{\alpha(x) \to \beta}{\exists x \alpha(x) \to \beta}$

(V) $\forall x(\alpha(x))^o \to (\forall x \alpha(x))^o$

(VI) $\forall x((\alpha(x))^o \to (\exists x \alpha(x))^o$

where the variables x and y and the formulas α and β satisfy the usual restrictions.

Furthermore, an additional postulate needs to be adopted. To present it, let us say that two formulas are *congruent* if one can be obtained from the other by replacing bound variables or by suppressing vacuous quantifications (without confusion of variables). Here is the additional postulate:

(VII) $A \leftrightarrow B$, where A and B are congruent formulas.

A reason to accept postulates I - IV derives from the fact that they hold in classical logic and are acceptable intuitively. Postulates V and VI are assumed for reasons similar to those that motivate the postulates of C_1. Finally, VII is required for it seems correct to say that two congruent formulas are equivalent, and this fact cannot be deduced from the remaining postulates.

We can then establish the following theorem, whose proof is done in the same way as in classical logic.

THEOREM 47. *All derived rules of theorem 5 hold in $C_1^\star$. Furthermore, we have:*

$\alpha(x) \vdash \forall x \alpha(x)$ $\qquad$ $\forall x \alpha(x) \vdash \alpha(y)$

$$\alpha(t) \vdash \exists x\alpha(x) \qquad \text{If } \Gamma \cup \{\alpha(x)\} \vdash \beta, \text{ then } \Gamma \cup \{\exists x\alpha(x)\} \vdash \beta,$$

with the usual restrictions (*see* [Kleene, 1952]).

Note that in $C_1^\star$, similarly to what happens in the classical calculus $C_0^\star$, the above rules impose restrictions on the deduction theorem (see [Kleene, 1952]).

The proofs of the following theorems are also obtained without difficulty.

THEOREM 48. *In* $C_1^\star$*:*

$$\vdash \forall x(\alpha(x) \to \beta(x)) \to (\forall\alpha(x) \to \forall x\beta(x))$$

$$\vdash \forall x(\alpha(x) \to \beta(x)) \to (\exists\alpha(x) \to \exists x\beta(x))$$

THEOREM 49. *Suppose that* $\gamma_1, \dots, \gamma_n$ *are the quantificationally prime components of the formula* α *and the set of formulas* Γ. *In this case, if* $\Gamma \vdash \alpha$ *in the classical predicate calculus, then* $\{\gamma_1^o, \dots, \gamma_n^o\} \cup \Gamma \vdash \alpha$ *in* $C_1^\star$, *and conversely.*

COROLLARY 50. *Let* $\alpha_1, \dots, \alpha_n$ *be the quantificationally prime components of the formula* α. *Then a necessary and sufficient condition for* $\vdash \alpha$ *in* $C_0^\star$ *is that* $\{\alpha_1^o, \dots, \alpha_n^o\} \vdash \alpha$ *in* $C_1^\star$.

THEOREM 51. $C_0^\star$ *is included in* $C_1^\star$.

Proof. In the case of formulas whose prime components are well-behaved, Theorem 49 shows that, for these formulas, the laws and rules of $C_0^\star$ hold. But, from Theorem 24, C_1 is, in a certain sense, included in $C_1^\star$, and so is $C_0^\star$. ■

COROLLARY 52. $C_1^\star$ *is undecidable.*

Proof. Given that $C_0^\star$ is included in $C_1^\star$, and $C_0^\star$ is undecidable (Church's theorem), so is $C_1^\star$. ■

THEOREM 53. *The following schemas are valid in* $C_1^\star$, *with the usual restrictions* (see [Kleene, 1952, p. 162]), *and where* γ *is a formula in which* x *does not appear free:*

1. $\forall x\gamma \leftrightarrow \gamma$
2. $\exists x\gamma \leftrightarrow \gamma$
3. $\forall x\forall y\alpha(x,y) \leftrightarrow \forall y\forall x\alpha(x,y)$
4. $\exists x\exists y\alpha(x,y) \leftrightarrow \exists y\exists x\alpha(x,y)$
5. $\forall x\forall y\alpha(x,y) \to \forall x\alpha(x,x)$
6. $\exists x\alpha(x,x) \to \exists x\exists y\alpha(x,y)$
7. $\forall x\alpha(x) \to \exists x\alpha(x)$
8. $\exists x\forall y\alpha(x,y) \to \forall y\exists x\alpha(x,y)$

9. $\forall x(\alpha(x) \wedge \beta(x)) \leftrightarrow \forall x\alpha(x) \wedge \forall x\beta(x)$

10. $\exists x(\alpha(x) \vee \beta(x)) \leftrightarrow \exists x\alpha(x) \vee \exists x\beta(x)$

11. $\alpha \wedge \forall x\beta(x) \leftrightarrow \forall x(\alpha \wedge \beta(x))$

12. $\alpha \vee \exists x\beta(x) \leftrightarrow \exists x(\alpha \vee \beta(x))$

13. $\alpha \wedge \exists x\beta(x) \leftrightarrow \exists x(\alpha \wedge \beta(x))$

14. $\alpha \vee \forall x\beta(x) \rightarrow \forall x(\alpha \vee \beta(x))$

15. $\exists x(\alpha(x) \wedge \beta(x)) \rightarrow \exists x\alpha(x) \wedge \exists x\beta(x)$

16. $\forall x\alpha(x) \vee \forall x\beta(x) \rightarrow \forall x(\alpha(x) \vee \beta(x))$

Proof. Similar to the proofs in the classical calculus (the same point also applies to some other schemas that hold in $C_1^\star$). ■

THEOREM 54. *In* $C_1^\star$,

$$\forall x(\alpha(x))^o \vdash \exists x\alpha(x) \leftrightarrow \neg\forall\ \neg\alpha(x)$$

$$\forall x(\alpha(x))^o \vdash \forall x\alpha(x) \leftrightarrow \neg\exists x\neg\alpha(x)$$

$$\forall x(\alpha(x))^o \vdash \neg\forall x\alpha(x) \leftrightarrow \exists x\neg\alpha(x)$$

$$\forall x(\alpha(x))^o \vdash \neg\exists x\alpha(x) \leftrightarrow \forall x\neg\alpha(x)$$

We define now the *k-transform* of a formula α, where k is a numeral among $\underline{1}, \underline{2}, \ldots$, that is, constant symbols in correspondence with the natural numbers $1, 2, \ldots$ [Kleene, 1952, pp. 177ff]. By hypothesis, these symbols do not belong to the language of $C_0^\star$ and of $C_1^\star$.

Case 1: If α has no free variables, its k-transform is obtained as follows: each part of α of the form $\forall x\beta(x)$ or of the form $\exists x\beta(x)$ is replaced respectively by $\beta(\underline{1}) \wedge \beta(\underline{2}) \wedge \ldots \wedge \beta(\underline{n})$ or by $\beta(\underline{1}) \vee \beta(\underline{2}) \vee \ldots \vee \beta(\underline{n})$. Hence, this k-transform has no (individual) variables.

Case 2: If α has free variables, that is, if α is $\alpha(x_1, \ldots, x_m)$, then its k-transforms are obtained as follows: (a) we replace the variables $x_1, \ldots, x_m$ by permutations of $1, 2, \ldots, k$ with repetitions of order k; (b) we then take the k-transforms of the resulting formulas of (a).

THEOREM 55. *If* $\Gamma \vdash \alpha$ *in* $C_1^\star$, *then any k-transform of* α *is deducible, in* C_1, *from the k-transforms of the formulas of* Γ.

Proof. By induction on the length of a deduction of α from Γ in the calculus $C_1^\star$. The proof is similar to the classical case (that is, similar to $C_0^\star$). ■

COROLLARY 56. *If* $\vdash \alpha$ *in* $C_1^\star$*, then the k-transforms of* α *are provable in* C_1*.*

COROLLARY 57. *If the formula* α *has only predicate symbols of arity zero, that is, with zero associated variables, and* $\vdash \alpha$ *in* $C_1^\star$*, then* $\vdash \alpha$ *in* C_1*.*

Proof. It suffices to note that, in this case, the only k-transform of α is α itself. ■

Remark: Corollary 57 is important for it emphasizes that those propositional schemas that do not hold in C_1 continue not to hold in $C_1^\star$. In other words, if we add to C_1 the postulates and specific formation rules of $C_1^\star$, no new result is obtained relatively to the *pure* formulas of C_1. Furthermore, Theorem 55 can be extended to sub-systems of the classical propositional calculus (and of intuitionistic propositional calculus) and their corresponding quantification theories. This is the case, for instance, of classical positive logic (and intuitionistic positive logic) as well as the propositional intuitionistic calculus and the minimal intuitionistic calculus. Theorem 55 applies to the classical calculi C_0 and $C_0^\star$ and constitutes the so-called Hilbert-Bernays Theorem on k-transforms.

THEOREM 58. *In* $C_1^\star$*, the following schemas are not valid:*

1. $\neg\exists x \neg\alpha(x) \leftrightarrow \forall x \alpha(x)$
2. $\neg\forall x \neg\alpha(x) \leftrightarrow \exists x \alpha(x)$
3. $\neg\exists x \alpha(x) \leftrightarrow \forall x \neg\alpha(x)$
4. $\exists x \neg\alpha(x) \leftrightarrow \neg\forall x \alpha(x)$

Proof. It is easy to see that the 2-transforms of these schemas are not provable in $C_1^\star$. It suffices to take the 2-transforms and to use the tables of Theorem 6. ■

THEOREM 59. $C_1^\star$ *is strictly weaker than* $C_0^\star$*, and* $C_0^\star$ *can be obtained from* $C_1^\star$ *by adding to its postulates the following schema:* $\neg(\alpha \wedge \neg\alpha)$*.*

Proof. The result is a consequence of the fact that C_0 is obtained from C_1 if we add $\neg(\alpha \wedge \neg\alpha)$ to C_1 as a postulate. ■

The remaining calculi $C_n^\star$, $2 \leq n < \omega$ of the hierarchy (4) are obtained by adding to C_n, with obvious adaptations, the postulates I-IV and VII above, plus the following ones:

$$(\mathrm{V}_n)\ \ \forall\, (\alpha(x))^{(n)} \rightarrow (\forall x \alpha(x))^{(n)}$$

$$(\mathrm{VI}_n)\ \forall\, (\alpha(x))^{(n)} \rightarrow (\exists x \alpha(x))^{(n)}$$

$C_\omega^\star$ is obtained from C_n by adding to its postulates the schemes IV and VII.

It is clear that the calculi $C_n^\star$, $2 \leq n < \omega$ have properties similar to those of $C_1^\star$. In particular, the following results hold.

THEOREM 60. *Let the quantificationally prime components of the formula* α *and of the formulas of* Γ *be* $\alpha_1, \ldots, \alpha_m$*. In this case, if* $\Gamma \vdash \alpha$ *in* $C_0^\star$*, then* $\alpha_1^{(n)}, \ldots \alpha_m^{(n)}, \Gamma \vdash \alpha$ *in* $C_n^\star$*,* $1 \leq n < \omega$*, and conversely.*

THEOREM 61. *$C_n^\star$, $0 \leq n < \omega$, is undecidable.*

THEOREM 62. *(Essenin-Volpin) $C_\omega^\star$ is undecidable.*

THEOREM 63. *If $\Gamma \vdash \alpha$ in $C_n^\star$, $0 \leq n \leq \omega$, then the k-transforms of α are deducible, in C_n, from the k-transforms of the formulas of Γ.*

THEOREM 64. *If $\vdash \alpha$ in $C_n^\star$, $0 \leq n \leq \omega$, then the k-transforms of α are provable in C_n.*

THEOREM 65. *If formula α belongs to C_n, $0 \leq n < \omega$, and $\vdash \alpha$ in $C_n^\star$, then $\vdash \alpha$ in C_n.*

THEOREM 66. *Every calculus of the hierarchy $C_n^\star$, $1 \leq n \leq \omega$, is a proper sub-system of $C_0^\star$.*

We can prove for the hierarchy (4) results analogous to those established for the corresponding propositional calculi. For instance, we have:

THEOREM 67. *Every calculus of the hierarchy $C_n^\star$, $0 \leq n \leq \omega$, is strictly stronger than those that follow it.*

THEOREM 68. *Every calculus $C_n^\star$, $0 \leq n < \omega$, is finitely trivializable.*

THEOREM 69. *$C_\omega^\star$ is not finitely trivializable.*

2.5 *Equality*

From the hierarchy (4), we can construct the following hierarchy of predicate calculi with equality:

(5) $C_0^=, C_1^=, C_2^=, \ldots, C_n^=, \ldots, C_\omega^=.$

This is done by adding to their languages the binary predicate symbol of equality, =, with suitable modifications in the concept of formula, and by adding the following two postulates:

(I') $\forall x(x = x)$

(II') $x = y \rightarrow (\alpha(x) \rightarrow \alpha(y))$

with the same restrictions adopted in the classical calculus. Here, $C_0^=$ stands for the classical first-order predicate calculus with equality. We begin by studying $C_1^=$.

THEOREM 70. *We have in $C_1^=$:*

1. $\vdash x = x$
2. $\vdash x = y \rightarrow y = x$
3. $\vdash x = y \wedge y = z \rightarrow x = z$

THEOREM 71. *We have in $C_1^=$:*

1. $\vdash x = y \leftrightarrow (\alpha(x) \rightarrow \alpha(y))$

2. $\forall t(\alpha(t))^o \vdash \alpha(x) \wedge \neg\alpha(y) \to x \neq y$, where $x \neq y$ stands for $\neg(x = y)$.

Proof. The first schema is proven in the same way as in $C_0^=$. As for the second, note that $\forall t(\alpha(t))^o \vdash (\alpha(y))^o$ and $x = y \to (\alpha(x) \to \alpha(y))$. So,

$$\forall t(\alpha(t))^o, \alpha(x) \wedge \neg\alpha(y), x = y \vdash (\alpha(y))^o$$

$$\forall t(\alpha(t))^o, \alpha(x) \wedge \neg\alpha(y), x = y \vdash \neg\alpha(y)$$

$$\forall t(\alpha(t))^o, \alpha(x) \wedge \neg\alpha(y), x = y \vdash \alpha(y).$$

Consequently,

$$\forall t(\alpha(t))^o, \alpha(x) \wedge \neg\alpha(y) \vdash \neg(x = y)$$

$$\forall t(\alpha(t))^o \vdash \alpha(x) \wedge \neg\alpha(y) \to x \neq y.$$

■

THEOREM 72. *In* $C_1^=$*:*

$$\forall y(\alpha(y) \leftrightarrow \exists x(x = y \wedge \alpha(x))$$

$$\forall y(\alpha(y) \leftrightarrow \forall x(x = y \to \alpha(x))$$

$$\forall y \exists x(x = y)$$

THEOREM 73. *If we add to the postulates of* $C_1^=$ *the schema* $\neg(\alpha \wedge \neg\alpha)$*, we obtain* $C_0^=$.

THEOREM 74. *Let* $\Gamma \vdash \alpha$*, and let* $\alpha_1, \ldots, \alpha_m$ *be the quantificationally prime components of the formula* α *and of the formulas in* Γ*. In this case,* $\Gamma \vdash \alpha$ *in* $C_0^=$ *iff* $\alpha_1^o, \ldots, \alpha_m^o, \Gamma \vdash \alpha$ *in* $C_1^=$.

THEOREM 75. $C_\omega^=$ *is undecidable.*

Proof. $C_1^=$ contains $C_0^=$ just as $C_1^\star$ contains $C_0^\star$. So, $C_1^=$ is undecidable. ■

All valid schemas of $C_0^=$ that do not explicitly contain the symbol $\neg$ are valid in $C_1^=$.

THEOREM 76. *If* $\Gamma \vdash \alpha$ *in* $C_0^=$*, and if in the formulas of* $\Gamma \cup \{\alpha\}$ *we replace* $\neg^\star$ *for* $\neg$*, thus obtaining the formula* $\alpha^\star$ *and the set of formulas* $\Gamma^\star$*, then* $\Gamma^\star \vdash \alpha^\star$ *in* $C_1^=$.

Proof. It suffices to note that $\neg^\star$ has the properties of classical negation. ■

DEFINITION 77. $\exists! x\alpha(x) =_{\mathrm{def}} \exists y \forall x(y = x \leftrightarrow \alpha(x))$

THEOREM 78. *In* $C_1^=$, $\vdash \forall x \exists! y(x = y)$.

THEOREM 79. *If* α *does not contain the equality symbol and* $\vdash \alpha$ *in* $C_1^=$*, then* $\vdash \alpha$ *in* $C_1^\star$.

Proof. Let D be a formal proof of α in $C_1^=$. There are only a finite number of applications of postulate (II') in D, say

$$x_1 = y_1 \to (\alpha_1(x_1) \leftrightarrow \alpha_1(y_1)), \ldots, x_k = y_k \to (\alpha_k(x_k) \leftrightarrow \alpha_k(y_k)).$$

Let u and v be variables that don't appear in the formulas of D. We denote by $K(u, v)$ the universal closure of the following formula, with respect to all free variables distinct from u and v:

$$(\alpha_1(x_1) \leftrightarrow \alpha_1(y_1)) \wedge \ldots \wedge (\alpha_k(x_k) \leftrightarrow \alpha_k(y_k)).$$

If we then replace subformulas of the form $x = y$ of formulas occurring in D by $K(x, y)$, we get a formal proof of α in $C_1^\star$, after a few suitable adaptations. ■

As a consequence of the above theorem, we have that if $\Gamma \cup \{\alpha\}$ is a set of formulas of $C_1^\star$ such that $\Gamma \vdash \alpha$ in $C_1^=$, then $\Gamma \vdash \alpha$ also in $C_1^\star$.

The theorem 79 is important, for it shows that the quantificational schemes that are not deducible in $C_1^\star$ remain not deducible in $C_1^=$. So, the formulas $\alpha \wedge \neg\alpha \to \beta$, $\neg(\alpha(x) \wedge \neg\alpha(x))$ and $\exists x\alpha(x) \leftrightarrow \neg\forall x\neg\alpha(x)$ are not theorems of $C_1^=$, since they are not theorems of $C_1^\star$.

With regard to the calculi of the hierarchy (5), we can prove several interesting results. In what follows, we sum up some of these results, without proof.

THEOREM 80. *The propositions* 70, 72, 73, 75 *and* 79 *hold for* $C_n^=$, $0 \leq n \leq \omega$.

THEOREM 81. *If the quantificationally prime components of the formulas of* $\Gamma \cup \{\alpha\}$ *are* $\alpha_1, \ldots, \alpha_m$, *then if* $\Gamma \vdash \alpha$ *in* $C_0^=$, *we have that* $\alpha_1^o, \ldots, \alpha_m^o, \Gamma \vdash \alpha$ *in* $C_n^=$, $1 \leq n \leq \omega$.

THEOREM 82. *Every calculus of the hierarchy (5) is strictly stronger than those following it.*

THEOREM 83. *The calculi of the hierarchy* $C_n^=$, $1 \leq n < \omega$, *is finitely trivializable, while* $C_\omega^=$ *is not.*

2.6 Descriptions

In the calculi $C_n^=$, $1 \leq n \leq \omega$, we can introduce the description operator, ι, by means of a contextual definition, similarly to Russell's well-known definition in the case of $C_0^=$. Alternatively, the symbol ι can be introduced in the language of the calculi as a primitive symbol, satisfying the axioms presented below. This gives rise to a new hierarchy of calculi of descriptions:

(6) $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n, \ldots, \mathcal{D}_\omega$.

The postulates to be added to those of $C_n^=$, $1 \leq n \leq \omega$, are the following ones (satisfying the usual restrictions; see [Rosser, 1953]):

(D.1) $\forall x F(x) \to F(\iota y Q(y))$

(D.2) $\forall x(P(x) \leftrightarrow Q(x)) \to \iota x P(x) = \iota x Q(x)$

(D.3) $\iota x F(x) = \iota y F(y)$

(D.4) $P(\iota y Q(y)) \to \exists x P(x)$

(D.5) $\exists! x P(x) \to (\forall x((\iota x P(x) = x) \leftrightarrow P(x))$.

It can be proven that $\mathcal{D}_n$ is a conservative extension of $C_n^=$ $(0 \leq n \leq \omega)$. As above, we will focus the discussion on the first calculus of the hierarchy, namely $\mathcal{D}_1$. In particular, the following results are significant (and their proofs are adaptations of the classical case):

THEOREM 84. *In* $\mathcal{D}_1$, $\vdash \iota x F(x) = \iota x F(x)$.

THEOREM 85. *If* $\iota x F(x)$ *is free for* x *in* $F(x)$, *then* $\vdash \exists! x F(x) \to F(\iota x F(x))$.

Proof. From (D.5), we have $\vdash \exists! x F(x) \to \forall x((\iota x F(x) = x) \leftrightarrow F(x))$. Taking $\exists! x F(x)$ as hypothesis, we get $\forall x((\iota x F(x) = x) \leftrightarrow F(x))$. Then, from (D.1), $(\iota x F(x) = \iota x F(x)) \leftrightarrow F(\iota x F(x))$, hence $F(\iota x F(x))$. So, $\vdash \exists! x F(x) \to F(\iota x F(x))$. ■

THEOREM 86. *In* $\mathcal{D}_1$, $\vdash \forall y(\iota x(x = y) = y)$.

THEOREM 87. *In* $\mathcal{D}_1$, *we have:*

1. $\vdash \forall y(\iota x P(x) = y \leftrightarrow y = \iota x P(x))$
2. $\vdash \forall y \forall x(\iota x P(x) = y \wedge y = z \to \iota x P(x) = z)$
3. $\vdash \forall x \forall z(x = \iota y Q(y) \wedge \iota y Q(y) = z \to x = z)$
4. $\vdash \forall z(\iota x P(x) = \iota y Q(y) \wedge \iota y Q(y) = z \to \iota x P(x) = z)$
5. $\vdash \forall y(\iota x P(x) = y \wedge y = \iota z R(z) \to \iota x P(x) = \iota z R(z))$
6. $\vdash \iota x P(x) = \iota y Q(y) \leftrightarrow \iota y Q(y) = \iota x P(x)$
7. $\vdash \iota x P(x) = \iota y Q(y) \wedge \iota y Q(y) = \iota z R(z) \to \iota x P(x) = \iota z R(z)$
8. $\vdash F(\iota x P(x)) \wedge \exists x(x = \iota x P(x) \to F(x))$
9. $\vdash F(\iota x P(x)) \leftrightarrow \forall x(x = \iota x P(x) \to F(x))$
10. $\vdash \iota x P(x) = \iota y Q(y) \wedge \iota z Q(z) = \iota t K(t) \to \iota x P(x) = \iota t K(t)$

THEOREM 88. *Let* $F(x)$ *be a formula, and* u *and* t *be two terms (variables or descriptions) free for* x *in* $F(x)$. *Then,*

$$\vdash u = v \to (F(u) \to F(v)).$$

THEOREM 89. *If* y *does not occur free in* $P(x)$ *and* $\iota x P(x)$ *is free for* x *in* $P(x)$, *then*

$$\vdash \exists! x P(x) \to (F(\iota x P(x)) \leftrightarrow \exists y(F(y) \wedge \forall x(x = y \leftrightarrow P(x)))).$$

Proof. Let us assume that $\exists! x P(x)$. So,

$$\forall x(\iota x P(x) = x \leftrightarrow P(x)).$$

Then, taking $F(\iota x P(x))$ as a hypothesis, we get

$$F(\iota x P(x)) \text{ and } \forall x(\iota x P(x) = x \leftrightarrow P(x)).$$

Hence,

$$F(\iota x P(x)) \wedge \forall x(x = \iota x P(x) \leftrightarrow P(x)),$$

and so,

$$\exists y(F(y) \wedge \forall x(x = y \leftrightarrow P(x)).$$

Consequently,

$$F(\iota x P(x)) \to \exists y(F(y) \wedge \forall x(x = y \leftrightarrow P(x))).$$

But, assuming that

$$\exists y(F(y) \wedge \forall x(x = y \leftrightarrow P(x)),$$

it results that

$$F(y) \wedge \forall x(x = y \leftrightarrow P(x)),$$

so

$$F(y) \text{ and } \forall x(x = y \leftrightarrow P(x)),$$

and hence we conclude that $y = \iota x P(x)$ and that $F(\iota x P(x))$. Then,

$$\exists y(F(y) \wedge \forall x(x = y \leftrightarrow P(x)) \to F(\iota x P(x)).$$

So,

$$\exists! x P(x) \to (F(\iota x P(x)) \leftrightarrow \exists y(F(y) \wedge \forall x(x = y \leftrightarrow P(x)))).$$

■

THEOREM 90. *Let* $A_1, \ldots, A_n$ *be the quantificationally prime components of the formula* F *and of the formulas in* Γ. *In this case, if* $\Gamma \vdash F$ *in* $\mathcal{D}_0$, *then* $A_1^o, \ldots, A_n^o, \Gamma \vdash F$ *in* $\mathcal{D}_1$, *and conversely.*

Let us consider the set of formulas Γ of $\mathcal{D}_1$, and suppose that the finitely many descriptions that appear in these formulas are:

$$\begin{array}{l} \iota x_1 Q_1(x_1), \iota x_2 Q_1(x_2), \ldots \\ \iota x_1 Q_2(x_1), \iota x_2 Q_2(x_2), \ldots \\ \vdots \end{array}$$

which we admit ordered by some order relation. We can then associate to each of these descriptions one of the variables $t_0, \ldots, t_{n-1}$, which do not appear in the formulas of Γ, in such a way that the association is a bijection. The ι-transforms of the formulas of Γ are the formulas obtained from the members of Γ by replacing in them the descriptions by their associated variables, and identifying the variables corresponding to descriptions, such as in the case of D.4.

THEOREM 91. *If* $\vdash F$ *in* $\mathcal{D}_1$ *and* F *does not contain the symbols* $=$ *and* ι, *then* $\vdash F$ *in* $C_1^\star$.

Proof. If F can be proven in $\mathcal{D}_1$, it is deducible in $C_1^{=\star}$ from a finite number of formulas of types D.1–D.5 (see page 817 above), where $C_1^{=\star}$ is the calculus obtained from $C_1^=$ by introducing formation rules enabling descriptions as new terms in addition to variables. Let us denote by Δ a deduction of F in $C_1^{=\star}$ from formulas of types D.1–D.5. If we substitute, in such a deduction, formulas by their ι-transforms, we see that we get a valid deduction $\Delta^\star$ in $C_1^=$. But after such substitutions, the formulas of types D.1, D.3 and D.4 turn out to be axioms of $C_1^=$ and $\Delta^\star$ is a deduction of F from formulas of the forms (i) $\forall x(P(x) \leftrightarrow Q(x)) \rightarrow t_i = t_j$ and (ii) $\exists! x P(x) \rightarrow \forall x(t_k = x \leftrightarrow P(x))$. If in $\Delta^\star$ we identify pairs of variables, such as t_i and t_j, that appear free in the formulas of the type (i) above, then $\Delta^\star$ becomes a deduction $\Delta^{\star\star}$, which is still valid in $C_1^=$. But since the formulas of the specified type are axioms of $C_1^=$, they can be suppressed from the hypothesis of $\Delta^{\star\star}$. This yields a deduction $\Delta^{\star\star\star}$ of F in $C_1^=$ from formulas of the form (ii). However, to say that F is deducible in $C_1^=$ from formulas of the above type is to say that F is deducible in $C_1^\star$ from formulas of the following forms: $\forall x(x = x)$, $x = y \rightarrow (A(x) \rightarrow A(y))$ and $P(t_k)$. Therefore, taking into account theorem 79, we easily complete the proof. ■

The calculi $\mathcal{D}_0, \mathcal{D}_2, \ldots, \mathcal{D}_n, \ldots, \mathcal{D}_\omega$ have properties similar to those of $\mathcal{D}_1$.

THEOREM 92. *The Theorems 84 to 89 hold for every calculi* $\mathcal{D}_n$, $0 \leq n \leq \omega$.

THEOREM 93. *If* $\vdash F$ *in* $\mathcal{D}_0$ *and* F *does not contain the symbol* ι, *then* $\vdash F$ *in* $C_0^=$.

THEOREM 94. *The calculi* $\mathcal{D}_n$, $0 \leq n \leq \omega$ *are simply consistent.*[7]

Proof. By the previous theorem, $\mathcal{D}_0$ is simply consistent if $C_0^=$ is. Since the remaining calculi are sub-systems of $\mathcal{D}_0$, the theorem follows. ■

THEOREM 95. *Let* $0 \leq n < \omega$, *and let* $A_1, \ldots, A_n$ *be the quantificationally prime components of the formula* F *and of the formulas in* Γ. *Then,* $\Gamma \vdash F$ *in* $\mathcal{D}_0$ *iff* $A_1^{(n)}, \ldots, A_n^{(n)}\Gamma \vdash F$ *in* $\mathcal{D}_0$.

THEOREM 96. *The calculi* $\mathcal{D}_n$, $0 \leq n < \omega$, *are undecidable.*

THEOREM 97. *Every calculus of the hierarchy* $\mathcal{D}_n$, $0 \leq n \leq \omega$, *is strictly stronger than those that follow it.*

THEOREM 98. $\mathcal{D}_n$, $0 \leq n < \omega$, *is finitely trivializable.*

THEOREM 99. $\mathcal{D}_\omega$ *is not finitely trivializable.*

---

[7] This means that for no formula A, both A and $\neg A$ are provable in these calculi ([Kleene, 1952, p. 124]).

THEOREM 100. *If* $\Gamma \vdash \alpha$ *in* $\mathcal{D}_n$ *and the formulas in* $\Gamma \cup \{\alpha\}$ *do not contain the description symbol, then* ‘$\Gamma \vdash \alpha$ *in* $C_n^=$, $0 \leq n \leq \omega$.

2.7 Semantics

We begin by sketching a two-valued semantics for C_1.[8] Here, $\mathcal{F}$ stands for the set of formulas of C_1; Γ and Δ designate subsets of $\mathcal{F}$, while $\overline{\Gamma}$ denotes the set of all formulas α such that $\Gamma \vdash \alpha$.

Recall that we say that a set Γ of formulas is trivial iff $\overline{\Gamma} = \mathcal{F}$; otherwise, it is nontrivial. Γ is inconsistent iff there is at least one formula α such that both α and $\neg\alpha$ belong to $\overline{\Gamma}$; otherwise, Γ is consistent. Finally, Γ is maximal nontrivial iff it is nontrivial and, for any formula α, if $\alpha \notin \Gamma$, then $\Gamma \cup \{\alpha\}$ is trivial.

THEOREM 101. *If* Γ *is maximal nontrivial, then:*

1. $\Gamma \vdash \alpha \Leftrightarrow \alpha \in \Gamma$
2. $\alpha \in \Gamma \Rightarrow \neg^\star\alpha \notin \Gamma$
3. $\neg^\star\alpha \in \Gamma \Rightarrow \alpha \notin \Gamma$
4. $\alpha \in \Gamma$ or $\neg^\star\alpha \in \Gamma$
5. $\vdash \alpha \Rightarrow \alpha \in \Gamma$
6. $\alpha, \alpha^o \in \Gamma \Rightarrow \neg\alpha \notin \Gamma$
7. $\neg\alpha, \alpha^o \in \Gamma \Rightarrow \alpha \notin \Gamma$
8. $\alpha, \alpha \rightarrow \beta \in \Gamma \Rightarrow \beta \in \Gamma$
9. $\alpha^o \in \Gamma \Rightarrow \alpha \notin \Gamma$ or $\neg\alpha \notin \Gamma$
10. $\alpha^o \in \Gamma \Rightarrow (\neg\alpha)^o \in \Gamma$

Proof. We will prove only one side of the first property. Suppose that $\Gamma \vdash \alpha$ but $\alpha \notin \Gamma$. Then, since Γ is maximal nontrivial, $\Gamma \cup \{\alpha\} \vdash \alpha \wedge \neg^\star\alpha$. Hence $\Gamma \vdash \alpha \rightarrow (\alpha \wedge \neg^\star\alpha)$, and $\Gamma \vdash \neg^\star\alpha$. But, since $\Gamma \vdash \alpha$, then $\Gamma \vdash \alpha \wedge \neg^\star\alpha$, therefore Γ is trivial, which is absurd. ■

DEFINITION 102. A *valuation* of C_1 is a mapping $v : \mathcal{F} \mapsto \{1, 0\}$ such that:

1) $v(\alpha) = 0 \Rightarrow v(\neg\alpha) = 1$
2) $v(\neg\neg\alpha) = 1 \Rightarrow v(\alpha) = 1$
3) $v(\beta^o) = v(\alpha \rightarrow \beta) = v(\alpha \rightarrow \neg\beta) = 1 \Rightarrow v(\alpha) = 0$
4) $v(\alpha \rightarrow \beta) = 1 \Leftrightarrow v(\alpha) = 0$ or $v(\beta) = 1$
5) $v(\alpha \wedge \beta) = 1 \Leftrightarrow v(\alpha) = v(\beta) = 1$
6), $v(\alpha \vee \beta) = 1 \leftrightarrow v(\alpha) = 1$ or $v(\beta) = 1$
7) $v(\alpha^o) = v(\beta^o) = 1 \Rightarrow v((\alpha \vee \beta)^o) = v(\alpha \wedge \beta)^o) = v((\alpha \rightarrow \beta)^o) = 1$.

THEOREM 103. *If* v *is a valuation of* C_1, *it has the following properties:* $v(\alpha) = 1 \Leftrightarrow v(\neg^\star\alpha) = 0$, $v(\alpha) = 0 \Leftrightarrow v(\neg^\star\alpha) = 1$, $v(\alpha) = 0 \Leftrightarrow v(\alpha) = 0$ *and* $v(\neg\alpha) = 1$, $v(\alpha) = 1 \Leftrightarrow v(\alpha) = 1$ *or* $v(\neg\alpha) = 0$.

DEFINITION 104. (i) A valuation v is *singular* if there exists at least one formula α such that $v(\alpha) = v(\neg\alpha) = 1$; otherwise, v is *normal*. (ii) A formula α is *valid* if for every valuation v, $v(\alpha) = 1$. (iii) A valuation v is a *model* for a set Γ of formulas if $v(\alpha) = 1$ for any $\alpha \in \Gamma$. (iv) A formula α is a *semantic consequence* of Γ if every model v of Γ is such that $v(\alpha) = 1$; in this case, we write $\Gamma \models \alpha$. In particular, $\models \alpha$ — which abbreviates $\emptyset \models \alpha$ — means that α is valid.

THEOREM 105 (Soundness). $\Gamma \vdash \alpha \Rightarrow \Gamma \models \alpha$ *(in particular,* $\vdash \alpha \Rightarrow \models \alpha$*).*

---

[8] The semantics of valuations, conceived as a general semantic method, was developed by the first author of this paper in the 1960s, in his Logic Seminar at Federal University of Paraná, Brazil.

Proof. As in classical logic. ■

LEMMA 106. *Every nontrivial set of formulas is contained in a maximal nontrivial set.*

Proof. By adapting the corresponding proof in classical logic. ■

COROLLARY 107. *There exist maximal nontrivial inconsistent sets.*

Proof. It is easy to see that $\{\alpha, \neg\alpha\}$ is inconsistent but nontrivial. By the preceding Lemma, it is contained in a maximal nontrivial set of formulas, which is inconsistent. ■

LEMMA 108. *Every maximal nontrivial set of formulas* Γ *has a model.*

Proof. Define a mapping $v : \mathcal{F} \mapsto \{0, 1\}$ as follows: for every formula α, if $\alpha \in \Gamma$, then $v(\alpha) = 1$, and $v(\alpha) = 0$ otherwise. It is then easy to see that v satisfies all the conditions in the definition of a valuation. ■

COROLLARY 109. *Any nontrivial set of formulas has a model.*

COROLLARY 110. *There are singular valuations; that is, valuations that assign value 1 to contradictory formulas. (A valuation that is not singular is called normal.)*

Proof. The set $\{\alpha, \neg\alpha\}$ is inconsistent but nontrivial, hence it is contained in a maximal nontrivial set, as shown above. But this set has a model, which is a singular valuation. ■

THEOREM 111 (Completeness). $\Gamma \models \alpha \Rightarrow \Gamma \vdash \alpha$ *(in particular,* $\models \alpha \Rightarrow \vdash \alpha$*).*

Proof. Let us suppose that $\Gamma \nvdash \alpha$. In this case, $\Gamma \cup \{\neg^\star\alpha\}$ is nontrivial and has a model v. So, there is a model v of Γ such that $v(\alpha) = 0$, which is absurd. ■

THEOREM 112. *There are inconsistent (but nontrivial) sets of formulas that have models.*

DEFINITION 113. Let $\Delta =_{\mathsf{def}} \{\alpha^o \in \mathcal{F} : \vdash \alpha\}$. Then Γ is said to be *strongly nontrivial* if $\Gamma \cup \Delta$ is nontrivial. Γ is said to be *strictly nontrivial* if $\Gamma \cup \Delta$ is nontrivial.

THEOREM 114. *There exist sets of formulas that are nontrivial and sets of formulas that are strictly nontrivial.*

Proof. To prove the first part of the theorem, suppose that Δ is the set $\{\alpha^o \in \mathcal{F} : \vdash \alpha\}$. Then Δ is consistent, which implies that it is also nontrivial. So, Δ is contained in a maximal nontrivial set Δ'. Let Δ'' be $\Delta' - \Delta$. Then Δ'' is strongly nontrivial. ■

We will now discuss a byproduct of the semantics for C_1 sketched above, namely, the decidability of this calculus.

The decidability of C_1.

DEFINITION 115 (Quasi-matrix). We call quasi-matrices the tables constructed according to the instructions given below, applicable to each formula of C_1(see [Alves, 1976]). Given a formula α:

1. Make a list of all propositional variables that appear in α, and arrange them in a line.

2. Under the propositional variables, place all possible distributions of 0s and 1s that can be attributed to the variables, as usual.

3. Make a list of the negations of propositional variables appearing in the formula. And calculate their truth-value, in each line, as follows: (i) if a variable has value 0, its negation will have truth-value 1; (ii) if the variable has value 1, bifurcate the line in which it appears, by writing in the first part the value 0 for the negation and, in the second part, the value 1 for the negation. Before a bifurcation, the truth-values must be the same for the two resulting lines.

4. Complete the previous list, by adding the sub-formulas of α and the negations of proper sub-formulas. And calculate, for each line, the truth-value of each sub-formula of α. If it is a proper sub-formula, calculate the value of its negation (whose proper sub-formulas and their negations have already been listed and calculated), as follows: (i) When no negations are involved, proceed as in the truth-tables for the classical propositional calculus; (ii) If any of the formulas in consideration is a negation — and so of the form $\neg\alpha'$ — write the truth-value 1 under it in those lines where α' has value 0. And in the lines where α' has value 1, proceed as follows: (i') If α' is of the form $\neg\beta$, then check if the value of β is the same as the value of $\neg\beta$; in this case, bifurcate the line by writing the value 0 in the first part and the value 1 in the second. If the value of β is distinct from the value of $\neg\beta$, simply write 0. (ii') If α' is of the form $\beta \bowtie \gamma$, where $\bowtie \in \{\rightarrow, \vee, \wedge\}$, there are two cases to consider: (a) α' is of the form $\delta \wedge \neg\delta$ or of the form $\neg\delta \wedge \delta$. In this case, write the value 0 for the formula α'; (b) α' is neither of the form $\delta \wedge \neg\delta$ nor of the form $\neg\delta \wedge \delta$. In this case, check if the value of β is equal to the value of $\neg\beta$ or if the value of γ is equal to the value of $\neg\gamma$. In the positive case, bifurcate the line, by writing the value 0 in the first part and, in the second, write 1. If the value of β is distinct from the value of $\neg\beta$, simply write 0.

We will exemplify this definition below. Before that, we state the following lemma:

LEMMA 116. $v : \mathcal{F} \mapsto \{0, 1\}$ *is a valuation iff:*

1. $v(\neg\alpha) = 0 \Rightarrow v(\alpha) = 1$

2. $v(\neg\neg\alpha) = 1 \Rightarrow v(\alpha) = 1$

3. $v(\beta^o) = v(\alpha \rightarrow \beta) = v(\alpha \rightarrow \neg\beta) = 1 \Rightarrow v(\alpha) = 0$

4. $v(\alpha \rightarrow \beta) = 1 \Leftrightarrow v(\alpha) = 0 \text{ or } v(\beta) = 1$

5. $v(\alpha \wedge \beta) = 1 \Leftrightarrow v(\alpha) = v(\beta) = 1$

6. $v(\alpha \vee \beta) = 1 \Leftrightarrow v(\alpha) = 1$ or $v(\beta) = 1$

7. $v((\alpha \wedge \beta)^o) = 0 \Leftrightarrow v(\alpha^o) = 0$ or $v(\beta^o) = 0$

8. $v((\alpha \vee \beta)^o) = 0 \Leftrightarrow v(\alpha^o) = 0$ and $v(\beta^o) = 0$

9. $v((\alpha \rightarrow \beta)^o) = 0 \Leftrightarrow v(\alpha^o) = 0$ or $v(\beta^o) = 1$

LEMMA 117. $v(\alpha^o) = 0 \Leftrightarrow v(\alpha) = v(\neg\alpha) = 1$

Proof. (a) $v(\alpha^o) = 0$ implies that $v(\alpha \wedge \neg\alpha) = 1$ and $v(\alpha) = v(\neg\alpha) = 1$. (b) Suppose that $v(\alpha) = v(\neg\alpha) = 1$. If $v(\alpha^o) = 1$, then $v(\alpha) = v(\neg\alpha) = v(\alpha^o) = 1$, that is, $v(\alpha) = v(\neg^\star\alpha) = 1$, and v would not be a valuation. Hence $v(\alpha^o) = 0$ and, therefore, $v(\alpha) = v(\neg\alpha) = 1$. ■

LEMMA 118. $v : \mathcal{F} \mapsto \{0, 1\}$ *is a valuation iff the conditions 1-6 of Lemma 116 hold and:*

7i. $v((\alpha \rightarrow \beta)^o) = 0 \Leftrightarrow v(\alpha) = v(\neg\alpha) = 1$ or $v(\beta) = v(\neg\beta) = 1$

7ii. $v((\alpha \wedge \beta)^o) = 0 \Leftrightarrow v(\alpha) = v(\neg\alpha) = 1$ or $v(\beta) = v(\neg\beta) = 1$

7ii. $v((\alpha \vee \beta)^o) = 0 \Leftrightarrow v(\alpha) = v(\neg\alpha) = 1$ or $v(\beta) = v(\neg\beta) = 1$

DEFINITION 119. Let v be a valuation and α a formula. Then v_α is called the *restriction* of v to the set of sub-formulas of α and the negations of proper sub-formulas of α.

LEMMA 120. *For every valuation v and formula α, $v(\alpha) = v_\alpha(\alpha)$.*

DEFINITION 121. Let v be a valuation and Γ a set of formulas. Then v_Γ is the *restriction* of v to the set Γ.

DEFINITION 122. We say that a line of a quasi-matrix *corresponds* to v_Γ if $v_\Gamma(\alpha)$ is the value corresponding to α in that line for every $\alpha \in \Gamma$, where Γ is the set of all formulas of the matrix.

LEMMA 123. *Given a quasi-matrix Q, then for every valuation v there exists a line of Q that corresponds to v_Γ, where Γ is the set of all formulas of Q.*

Proof. By induction on the number of columns of Q. ■

DEFINITION 124. Let Q be a quasi-matrix for a formula α, and let Γ be the set of all sub-formulas and negations of proper sub-formulas of α. Let k be a line of Q and $k(\alpha)$ be the value attributed to α in k. We call $\Delta(\Gamma, k)$ the set of formulas such that, for every formula α:

(I) *If $\alpha \in \Gamma$, then $\alpha \in \Delta(\Gamma, k)$ iff $k(\alpha) = 0$.*

(II) *If $\alpha \notin \Gamma$, then $\alpha \in \Delta(\Gamma, k)$ iff*

a) α is atomic, or

b) α is of the form $\neg\alpha_1$ and $\alpha_1 \notin \Delta(\Gamma, k)$, or *c)* α *is* $\alpha_1 \wedge \alpha_2$ *and* $\alpha_1 \in \Delta(\Gamma, k)$ *and* $\alpha_2 \in \Delta(\Gamma, k)$*, or*

d) α *is* $\alpha_1 \vee \alpha_2$ *and* $\alpha_1 \in \Delta(\Gamma, k)$ *and* $\alpha_2 \in \Delta(\Gamma, k)$*, or*

e) α *is* $\alpha_1 \to \alpha_2$ *and* $\alpha_1 \notin \Delta(\Gamma, k)$ *and* $\alpha_2 \in \Delta(\Gamma, k)$.

Some properties of the set $\Delta(\Gamma, k)$ are the following:

1. $\neg\alpha \in \Delta(\Gamma, k) \Rightarrow \alpha \notin \Delta(\Gamma, k)$

2. $\alpha \in \Delta(\Gamma, k) \Rightarrow \neg\neg\alpha \in \Delta(\Gamma, k)$

3. $\neg^{\star}\alpha \in \Delta(\Gamma, k) \Leftrightarrow \alpha \notin \Delta(\Gamma, k)$

4. $\alpha \to \beta \notin \Delta(\Gamma, k) \Leftrightarrow \alpha \in \Delta(\Gamma, k) \, \text{or} \beta \notin \Delta(\Gamma, k)$

5. $\alpha \in \Delta(\Gamma, k) \, \text{or} \beta \in \Delta(\Gamma, k) \Leftrightarrow \alpha \wedge \beta \in \Delta(\Gamma, k)$

6. $\alpha \notin \Delta(\Gamma, k) \, \text{or} \beta \notin \Delta(\Gamma, k) \Leftrightarrow \alpha \vee \beta \notin \Delta(\Gamma, k)$

7. $(\alpha \bowtie \beta)^o \in \Delta(\Gamma, k) \Rightarrow \alpha \notin \Delta(\Gamma, k) \, \text{and} \neg\alpha \notin \Delta(\Gamma, k)$, $\text{or} \beta \notin \Delta(\Gamma, k) \, \text{and} \neg\beta \notin \Delta(\Gamma, k)$, where $\bowtie \in \{\to, \wedge, \vee\}$.

LEMMA 125 (A. Loparić). *For every line k of a quasi-matrix Q, there is a valuation* v *such that* v_Γ *corresponds to k, where* Γ *is the set of formulas of Q.*

Proof. Let v be a function from $\mathcal{F}$ in $\{0, 1\}$ such that, for every $\alpha \in \mathcal{F}$, $v(\alpha) = 0$ if $\alpha \in \Delta(\Gamma, k)$, and $v(\alpha) = 1$ if $\alpha \notin \Delta(\Gamma, k)$. Then, by the properties 1-7 above of the set $\Delta(\Gamma, k)$, v is a valuation. Since v_Γ and k are the same, we can say that there exists a valuation v such that v_Γ corresponds to k. ■

THEOREM 126 (M. Fidel). *The calculus* C_1 *is decidable.*

Proof. Consequence of Lemmas 120, 123 and 125. The formula α is a theorem of C_1 iff in any quasi-matrix for α, the last column has only 1's. In effect, in this case, for every valuation v, we have that $v(\alpha) = v_\alpha(\alpha) = 1$. ■

Let us exemplify the method presented here by showing that $\neg(\alpha \vee \beta) \to \neg\alpha \wedge \neg\beta$ is not valid in C_1:

α	β	$\neg\alpha$	$\neg\beta$	$\alpha \vee \beta$	$\neg(\alpha \vee \beta)$	$\neg\alpha \wedge \neg\beta$	$\neg(\alpha \vee \beta) \to \neg\alpha \wedge \neg\beta$
0	0	1	1	0	1	1	1
1	0	0	1	1	0	0	1
		1	1	1	0	1	1
					1	1	1
0	1	1	0	1	0	0	1
			1	1	0	1	1
					1	1	1
1	1	0	0	1	0	0	1
			1	1	0	0	1
					1	0	0
		1	0	1	0	0	1
					1	0	0
			1	1	0	1	1
					1	1	1

The extension of the semantics of C_1 to the systems $C_n, 2 \leq n < \omega$, is immediate. All definitions and theorems are the same, with obvious adaptations (for instance, $\neg^\star$ becomes $\neg^{(n)}$, and α^o becomes $\alpha^{(n)}$). When constructing quasi-matrices, if α is of the form $\beta^{(n-1)} \wedge \neg\beta^{(n-1)}$, or of the form $\neg\beta^{(n-1)} \wedge \beta^{(n-1)}$, we must write 0 for the formula α. We must have $v(\neg\alpha) = 0$, otherwise, due to clause 7 of Definition 102, $v(\beta \wedge \neg^{(n)}\beta) = 1$, and then v would not be a valuation. Note that this clause is precisely the one that characterizes the quasi-matrices of the systems $C_n, 1 \leq n < \omega$.

As an additional example, to see that the schema $(\alpha^{(n-1)} \wedge \neg\alpha^{(n-1)})^{(n)}$ is valid in C_n, but not in C_m, for $m > n$, it suffices to show that the schema $(\alpha \wedge \neg\alpha)^o$ is valid in C_1, but not in C_2. In fact, we have in C_1:

α	$\neg\alpha$	$\alpha \wedge \neg\alpha$	$\neg(\alpha \wedge \neg\alpha)$	$(\alpha \wedge \neg\alpha) \wedge \neg(\alpha \wedge \neg\alpha)$	$\neg((\alpha \wedge \neg\alpha) \wedge \neg(\alpha \wedge \neg\alpha))$
0	1	0	1	0	1
1	0	0	1	0	1
	1	1	0	0	1

However, in C_2, we have:

α	$\neg\alpha$	$\alpha \wedge \neg\alpha$	$\neg(\alpha \wedge \neg\alpha)$	$(\alpha \wedge \neg\alpha) \wedge \neg(\alpha \wedge \neg\alpha)$	$\neg((\alpha \wedge \neg\alpha) \wedge \neg(\alpha \wedge \neg\alpha))$
0	1	0	1	0	1
1	0	0	1	0	1
	1	1	0	0	1
			1	1	0
					1

It is also possible to show that C_ω is decidable; see [Loparić and Alves, 1980] and section 6.

Semantics for $C_1^=$

The semantics sketched above can be extended to the quantificational calculi described earlier. Let us begin by defining an *interpretation* for $C_1^=$ (in fact, for the language of $C_1^=$) as a pair $I = \langle D, \rho \rangle$, where D is a non-empty set and ρ is a mapping such that: (i) ρ associates an element $\rho(a) \in D$ to each individual constant a of the language of $C_1^=$; (ii) for each n-ary functional symbol f of the language of $C_1^=$, ρ associates an n-ary function from D^n to D, and (iii) to each n-ary predicate symbol P, other than identity, ρ associates an n-ary relation on D; (iv) to the identity symbol $=$, ρ associates the diagonal of D, namely, the set $\Delta_D =_{\mathsf{def}} \{\langle x, x \rangle : x \in D\}$.[9]

If I is an interpretation for $C_1^=$, then for each element of D, we choose a new individual constant c, the name of the element (as usual, different names are chosen for different elements). This new language, as is well known, is called the *diagram language* of $C_1^=$ relatively to I ([Shoenfield, 1967, p. 18]), and we refer to it as $DC_1^=$.

If I is an interpretation for $C_1^=$, a *valuation* v of $C_1^=$ is a map from the set of sentences of $DC_1^=$ in $\{0, 1\}$, defined as follows:

(1–7) Clauses 1-7 of Definition 102

8) $v(\forall x \alpha(x)) = 1$ iff for any individual constant c of $DC_1^=$, $v(\alpha(c)) = 1$

9) $v(\exists x \alpha(x)) = 1$ iff there exists a constant c of $DC_1^=$ so that $v(\alpha(c)) = 1$

10) $v(\forall x (\alpha(x))^o) = 1 \Rightarrow v((\forall x \alpha(x))^o) = v((\exists x \alpha(x))^o) = 1$

11) If α and β are congruent, then $v(\alpha) = v(\beta)$

12) $v(c = c') = 1$ iff $\rho(c) = \rho(c')$

13) $v(c = c') = 1$ and $v(\alpha(c)) = 1$, then $v(\alpha(c')) = 1$

We say that a valuation v *satisfies* a sentence α of $DC_1^=$ (and, in particular, of $C_1^=$) if $v(\alpha) = 1$. v is a *model* for a set Γ of sentences if $v(\alpha) = 1$ for every $\alpha \in \Gamma$, and α is a *semantic consequence* of Γ (in symbols, $\Gamma \vDash \alpha$) if $v(\alpha) = 1$ for every model v of Γ.

THEOREM 127 (Soundness). $\Gamma \vdash \alpha \Rightarrow \Gamma \vDash \alpha$, *where* $\Gamma \cup \{\alpha\}$ *is a set of sentences.*

Proof. As in the classical case, using induction on the length of the proof of α from Γ. ■

We say that Γ is *trivial* if $\Gamma \vdash \alpha$, for every α; otherwise, Γ is *nontrivial.* Γ is *inconsistent* if there exists α such that $\Gamma \vdash \alpha$ and $\Gamma \vdash \neg\alpha$; otherwise, Γ is *consistent.* Finally, a set Γ os sentences is a *Henkin set* if for any formula $\alpha(x)$ with just one free variable, there exists a constant c of the language of $C_1^=$ such that $\Gamma \vdash \exists x \alpha(x) \to \alpha(c)$.

THEOREM 128. *If* Γ *is non-trival (and a Henkin set), then it is contained in a maximal nontrivial (and also a Henkin set) set of sentences.*

Proof. As in the classical case. ■

---

[9] As we noted earlier, the language may contain neither individual constants nor functional symbols.

THEOREM 129. *If* Γ *is a maximal nontrivial Henkin set of sentences, then:*

(1) $\alpha \to \beta \in \Gamma \Leftrightarrow \alpha \notin \Gamma$ *or* $\beta \in \Gamma$ (2) $\alpha \wedge \beta \in \Gamma \Leftrightarrow \alpha, \beta \in \Gamma$
(3) $\alpha \vee \beta \in \Gamma \Leftrightarrow \alpha \in \Gamma$ *or* $\beta \in \Gamma$ (4) $\Gamma \vdash \alpha \Leftrightarrow \alpha \in \Gamma$
(5) $\alpha \in \gamma \Leftrightarrow \neg^\star \alpha \notin \Gamma$ (6) $\neg\alpha, \alpha^o \in \Gamma \Rightarrow \alpha \notin \Gamma$
(7) $\alpha, \alpha^o \in \Gamma \Rightarrow \neg\alpha \notin \Gamma$ (8) $\alpha \in \Gamma$ *or* $\neg^\star \alpha \in \Gamma$
(9) $\alpha, \alpha \to \beta \in \Gamma \Rightarrow \beta \in \Gamma$ (10) $\alpha^o \in \Gamma \Rightarrow (\neg\alpha)^o \in \Gamma$
(11) $\alpha^o, \beta^o \in \Gamma \Rightarrow (\alpha \to \beta)^o, (\alpha \wedge \beta)^o, (\alpha \vee \beta)^o \in \Gamma$
(12) $\forall x \alpha(x) \in \Gamma$ *iff for any constant c of* $C_1^=$, $\alpha(c) \in \Gamma$
(13) $\exists x \alpha(x) \in \Gamma$ *iff for some constant c of* $C_1^=$, $\alpha(c) \in \Gamma$
(14) $\forall x(\alpha(x))^o \in \Gamma \Rightarrow (\forall x \alpha(x))^o \in \Gamma$ *and* $(\exists x \alpha(x))^o \in \Gamma$
(15) *If* α *and* β *are congruent, then* $\alpha \leftrightarrow \beta \in \Gamma$
(16) $c = c', \alpha(c) \in \Gamma \Rightarrow \alpha(c') \in \Gamma$

Proof. As in the classical case, by using the strong negation $\neg^\star$ instead of the weak negation $\neg$. ■

THEOREM 130. *If* Γ *is a Henkin nontrivial set of sentences (either consistent or inconsistent), then* Γ *has a model.*

Proof. Consequence of the preceding theorem. ■

COROLLARY 131. *Every nontrivial set of sentences of the language of* $C_1^=$ *has a model.*

THEOREM 132 (Completeness). $\Gamma \models \alpha \Rightarrow \Gamma \vdash \alpha$

THEOREM 133 (Löwenheim-Skolem). *If* Γ *has an infinite model, then it has an infinite denumerable model.*

Other results can also be obtained, so that it is possible to construct a paraconsistent model theory. For instance, E. A. Alves has developed such a theory for a variant of the calculus C_1 to which he introduced an additional axiom, namely, $\neg\neg\alpha \leftrightarrow \alpha$ (see [Alves, 1984]). (Recall that $\alpha \to \neg\neg\alpha$ is not a theorem of C_1; see Theorem 6.) The preceding results can also be adapted in order to be applied to the calculi C_n, $1 \leq n < \omega$, as well as to the calculi with descriptions $\mathcal{D}_n$, $1 \leq n < \omega$.

2.8 Syllogism and paraconsistency

Similarly to the case of traditional syllogistic, which was interpreted within classical monadic predicate calculus, it is possible to develop a paraconsistent syllogistic. It is based, for instance, on the monadic calculus corresponding to the paraconsistent predicate logic $C_1^\star$. In order to do that, it suffices to translate the propositions A, I, E and O into $C_1^\star$. The translations (based on the classical case) are as follows:

Aab	$\forall x(a(x) \to b(x))$
Iab	$\exists x(a(x) \wedge b(x))$
Eab	$\forall x(a(x) \to \neg b(x))$

Oab $\exists x(a(x) \wedge \neg b(x))$

Two remarks are in order here. (1) The valid positive deductions in $C_0^\star$, the classical predicate calculus, are also valid in $C_1^\star$; that is, when no explicit negation is involved, the positive deductions of $C_0^\star$ and C_1^* are the same. (2) In $C_1^\star$, one can find paraconsistent predicates, for which there are objects that satisfy these predicates and, at the same time, do not satisfy them. In other words, for some predicate p, the following holds:

$$\exists x(p(x) \wedge \neg p(x)).$$

Thus, based on arguments similar to the ones found in the classical case, it is possible to verify the validity of inferences, and one changes accordingly the theories of opposition, conversion, immediate inferences and syllogism. (Each predicate in the universe of discourse has three parts: of the elements that satisfy it, of those that do not satisfy it, and of those that simultaneously satisfy it and do not satisfy it. Simple graphics supply then evidence for the validity, or for the invalidity, of certain inferences and conversions.)

Based on this approach, one can prove the following result. In the paraconsistent logic $C_1^\star$, all modes of the first and of the third figures of the syllogism are valid; none of the second is valid; and of the fourth, just Bramantip and Dimaris modes are valid. It is worth mentioning that since $C_1^\star$ has a strong negation, of a classical trend, and if such negation is adopted in the interpretation of syllogistic reasoning, then the classical theory is obtained. As is known, Łukasiewicz has axiomatized the theory of categorical syllogism, based on the classical propositional calculus and admitting as specific axioms certain categorical propositions, as well as some appropriate definitions. Based on the paraconsistent propositional calculus, for instance the calculus C_1, it is also possible to formulate an axiomatics for paraconsistent syllogistic, articulated in parallel lines to the theory just outlined. Moreover, we should note that there are further extensions or modifications of the Aristotelian syllogistic that also admit paraconsistent versions, such as Hamilton's, De Morgan's and Gergone's.

Aristotle's syllogistic and paraconsistency

A surprising result can be sketched as follows. The theory of syllogism can be developed in an appropriate language so that the only formulas turn out to be expressions corresponding to the categorical propositions A, E, I and O. We can then interpret the valid syllogisms (given by the four figures of Aristotelian syllogistic) as providing inference rules, and adopt the Aristotelian definition of contradictory propositions to mean contradictory formulas of our language. Thus, it is easy to see that, based on such a language, we can define contradictory theories (that is, those which involve contradictory formulas in the above sense), which are nontrivial (in the sense that not all formulas are theorems). This means that, with a suitable interpretation, Aristotle's syllogistic can be seen as a paraconsistent logic, since from two contradictory premisses we cannot deduce any proposition whatsoever. Analogously, Aristotle's modal syllogistic constitutes also a paraconsistent modal logic.

3 PARACONSISTENT SET THEORY

Cantor's naive set theory was based mainly on two fundamental principles: the postulate of extensionality (if the sets x and y have the same elements, then they are equal), and the postulate of comprehension or separation (every property determines a set, composed of the objects that have this property). The latter postulate, in the standard (first-order) language of set theory, becomes the following formula (or schema of formulas):

$$(7) \quad \exists y \forall x (x \in y \leftrightarrow F(x)).$$

Now, it is enough to replaces the formula $F(x)$ in (7) by $x \notin x$ to derive Russell's paradox. That is, the principle of comprehension (7) entails an inconsistency. Thus, if one adds (7) to first-order logic, conceived as the logic of a set-theoretic language, a trivial theory is obtained. There are also other paradoxes, such as those of Curry and Moh Schaw-Kwei, that indicate that (7) is trivial or, more precisely, trivializes set theory if its underlying logic is classical, even ignoring negation.[10] In other words, classical positive logic is incompatible with (7); the same holds also for several other logics, such as the intuitionistic one.

Classical set theories are distinguished by the restrictions they impose on (7), so that the paradoxes are avoided. In order that the theory thus obtained does not become too weak, some further axioms, besides extensionality and comprehension (with appropriate restrictions), are added, depending on the particular case. Thus, for instance, in Zermelo-Fraenkel (ZF), comprehension is formulated in the following way:

$$(8) \quad \forall z \exists y \forall x (x \in z \leftrightarrow (x \in \ \wedge F(x))),$$

where the variables are subject to the usual conditions. In ZF, then, $F(x)$ determines the subset of the elements of the set z that have the property F (or satisfy the formula $F(x)$). In the Kelly-Morse system, on the other hand, comprehension is formulated as follows:

$$(9) \quad \exists y \forall x (x \in y \leftrightarrow (F(x) \wedge \exists z (x \in z)).$$

In Quine's system NF, in turn, the notion of stratification is employed, and the schema of comprehension is written like (7), provided that $F(x)$ is stratified (and the standard conditions regarding the variables are met).

However, we can ask whether it is possible to examine the problem from a different viewpoint: what is needed to maintain the schema (7) *without* restrictions (disregarding the conditions on the variables)? The answer is immediate: one should change the underlying logic, so that (7) does not inevitably lead to trivialization. The comprehension schema, without strong restrictions, leads to contradictions. Hence, such a logic has to be paraconsistent. It was slowly verified that there are infinitely many ways of weakening the classical restrictions imposed on the comprehension schema, each of them corresponding to distinct categories of paraconsistent logic. Furthermore, extremely weak logics have been formulated as the underlying logics of set theories in which (7) is used without any restrictions (not related to the choice of variables).

---

[10] Let us exemplify this fact with Curry's paradox. In (7), substitute for $F(x)$ the expression $x \in x \to \alpha$, where α is an arbitrary formula in which y does not appear free. Then, by (7), we get $\exists y \forall x (x \in y \leftrightarrow (x \in x \to \alpha))$. In honor of Curry, let us call this set y, c. In this case, $\forall x (x \in c \leftrightarrow (x \in x \to \alpha))$. But then $c \in c \leftrightarrow (c \in c \to \alpha)$, hence (i) $c \in c \to (c \in c \to \alpha)$ and (ii) $(c \in c \to \alpha) \to c \in c)$. However, the law of contraction, $(\gamma \to (\gamma \to \beta)) \to (\gamma \to \beta)$, holds. So, (i) entails (iii): $c \in c \to \alpha$. From (ii) and (iii), we get $c \in c$. Finally, from this last sentence and from (iii), by *modus pones*, we obtain α. This shows that (7) entails triviality even without negation.

An important point is that several paraconsistent set theories contain the classical one, in Zermelo-Fraenkel's, Kelly-Morse's or Quine's formulations. Hence, paraconsistency goes beyond the classical domain, and allows for, among other things, the reconstruction of traditional mathematics (see [da Costa *et al.*, 1998a], [da Costa, 1986a], [da Costa *et al.*, 2004], [Mortensen, 1995]). It is fair to claim that paraconsistent theories extend the classical ones, just as Poncelet's imaginary geometry comprises standard Euclidean geometry.

Moreover, these considerations suggest a difficulty in the foundations of logic. Classical elementary logic (in fact, it is enough to consider only positive logic) and the comprehension postulate seem to be both evident. We may even claim that they are equally evident, or intuitive. However, they are mutually incompatible. This constitutes, thus, an intriguing case of incompatible evidences, which generates, in particular, a difficulty for classical logic.

Without presenting a detailed philosophical analysis, we just note that classical and paraconsistent theories develop different approaches to the handling of inconsistencies. Interestingly, the paraconsistent approach contributes to a better understanding of the classical position itself (by helping, e.g., to clarify the benefits and limitations the position has). The paraconsistent approach also provides a clearer understanding of negation (e.g., by distinguishing different meanings of this notion), and the approach articulates a more realistic perception of the possibility of discourse even if the principle of noncontradiction is partially left behind.

3.1 The systems $\mathcal{NF}_n$, $1 \leq n \leq \omega$

Here we begin by describing $\mathcal{NF}_1$.[11] The underlying logic of $\mathcal{NF}_1$ is the calculus $C_1^=$, that is, we assume the language of this calculus plus the propositional postulates $\to_1$ to $\neg_4$ (see page 799), the postulates (I)-(VII) for the predicate calculus (see page 811), and the postulates for equality (I') and (II') (see page 815). The specific postulates of $\mathcal{NF}_1$ are:

(NF1) (Extensionality) $\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \to x = y)$

(NF2) (Comprehension) $\exists y \forall x (x \in y \leftrightarrow \alpha(x))$, where x and y are distinct variables, y does not occur free in $\alpha(x)$ and this formula is either stratified or is of the form $x \notin x$.[12]

---

[11] Caiero and de Souza have developed a paraconsistent version of the *ML* system in [Caiero and de Souza, 1997].

[12] We recall that a formula is stratified if it is possible to replace each variable occurring in it by a numeral in the following manner: we replace everywhere the same variable by the same numeral so that, for each occurrence of $\in$, the numeral immediately following $\in$ is the immediate successor of the numeral immediately preceding $\in$ [Hatcher, 1982, p. 213].

If we add to $C_1^=$ the schema $\neg(\alpha \wedge \neg\alpha)$, we get the classical first-order predicate calculus with identity. The system NF of Quine is obtained by adding to this calculus the postulates (NF1) and (NF2) above, provided that the latter postulate meets the condition that $\alpha(x)$ be stratified. We denote Quine's system by $\mathcal{NF}_0$.

In order to introduce the set theories $\mathcal{NF}_n$, $1 \leq n < \omega$, we employ the calculus $C_n^=$, $1 \leq n < \omega$, as their underlying logics plus the specific postulates above. Several results can now be established:

THEOREM 134. *$\mathcal{NF}_n$, $1 \leq n < \omega$, contains $\mathcal{NF}_0$.*

Proof. It follows from the fact that if α is a theorem of $\mathcal{NF}_0$, and if we replace all occurrences of $\neg$ in this formula by $\neg^{(n)}$, obtaining α', then α' is provable in $\mathcal{NF}_n$, $1 \leq n < \omega$, where $\neg^{(n)}\alpha$ is the formula $\neg\alpha \wedge \alpha^{(n)}$. ■

THEOREM 135. *If $\mathcal{NF}_n$, $1 \leq n < \omega$, is nontrivial, then $\mathcal{NF}_0$ is consistent.*

Proof. Let us suppose that $\mathcal{NF}_n$, $1 \leq n < \omega$, is nontrivial and that $\mathcal{NF}_0$ is inconsistent. Let $\alpha_1, \ldots, \alpha_n$ be a proof of a contradiction in $\mathcal{NF}_0$, where α_n is $\beta \wedge \neg\beta$. Then, if $\alpha'_1, \ldots, \alpha'_n$ are formulas obtained from the α_i as explained in the proof of the previous theorem, then this last sequence would be a derivation of $\beta' \wedge \neg^{(n)}\beta'$ in $\mathcal{NF}_n$. But in this system $(\gamma \wedge \neg_{(n)}\gamma) \to \delta$ is a valid scheme, and so $\mathcal{NF}_n$ would be trivial. ■

We can also prove the following results:

THEOREM 136.

1. In the hierarchy $\mathcal{NF}_n$, $1 \leq n < \omega$, every system is stronger that those that follow it.
2. If $\mathcal{NF}_1$ is nontrivial, then all $\mathcal{NF}_n$, $1 < n < \omega$, are also nontrivial.
3. $\mathcal{NF}_n$, $1 \leq n < \omega$, is inconsistent.

Let us comment on item 3. In $\mathcal{NF}_n$, Russell set, that is, the set $R =_{\text{def}} \{x : x \notin x\}$, does exist (see the next subsection). In other words, in these systems we have: $\vdash \exists y \forall x(x \in y \leftrightarrow x \notin x)$. As we will see below, it is easy to prove that $R \in R \wedge R \notin R$. So, the systems $\mathcal{NF}_n$ are inconsistent. The other two items can be proved without difficulty (see [da Costa, 1986a]).

The next step is to show that if $\mathcal{NF}_0$ is consistent, then $\mathcal{NF}_1$ is nontrivial. Therefore, due to item 2 of the previous theorem, the consistency of $\mathcal{NF}_0$ entails the nontriviality of $\mathcal{NF}_n$, $1 \leq n < \omega$.

Let us first define a system $\mathcal{NF}_1^*$ as follows: We keep only the following propositional postulates (see page 799): $\to_1 - \to_3$, $\wedge_1 - \wedge_3$, $\vee_1 - \vee_3$, and add the following new postulates: $(\to_4^*)$ (Peirce's Law) $((\alpha \to \beta) \to \alpha) \to \alpha$, and $(\to_5^*)$ $(\neg\alpha \to \beta) \to ((\neg\alpha \to \neg\beta) \to \alpha)$, where β is not atomic. This new set of postulates provides, of course, an axiomatization for the classical propositional logic. The remaining postulates of $\mathcal{NF}_1^*$ are those of $\mathcal{NF}_1$, except for those that turn out to be redundant. For instance, since α^o is provable in $\mathcal{NF}_1^*$ when α is not atomic, it results that postulates (V) and (VI) (see page 811) are provable. So, $\mathcal{NF}_1$ is weaker than $\mathcal{NF}_1^*$.

LEMMA 137. *The consistency of $\mathcal{NF}_0$ entails the nontriviality of $\mathcal{NF}_1^*$.*

Proof. Let $\mathcal{V} =_{\mathsf{def}} \{x : x = x\}$, and let f be a map whose domain is the set of formulas of $\mathcal{NF}_1^*$ and whose range is the set of formulas of $\mathcal{NF}_0$, defined as follows:

$$\begin{array}{ll} 1. f(x = y) =_{\mathsf{def}} x = y & 2. f(x \in y) =_{\mathsf{def}} x \in y \\ 3. f(x \in y) =_{\mathsf{def}} x \in y & 4. f(x \notin y) =_{\mathsf{def}} x \in \mathcal{V} \wedge y \in \mathcal{V} \\ 5. f(\forall x \alpha) =_{\mathsf{def}} \forall x f(\alpha) & 6. f(\exists x \alpha) =_{\mathsf{def}} \exists x f(\alpha) \\ 7. f(\alpha \wedge \beta) =_{\mathsf{def}} f(\alpha) \wedge f(\beta) & 8. f(\alpha \vee \beta) =_{\mathsf{def}} f(\alpha) \vee f(\beta) \\ 9. f(\alpha \to \beta) =_{\mathsf{def}} f(\alpha) \to f(\beta) & \end{array} .$$

Then, using the preceding results, we can see that if α is a theorem of $\mathcal{NF}_0$, then $f(\alpha)$ is a theorem of $\mathcal{NF}_1^*$. Since the rules of inference of $\mathcal{NF}_1^*$ are valid in $\mathcal{NF}_0$, any theorem α of $\mathcal{NF}_1^*$ induces a theorem $f(\alpha)$ of $\mathcal{NF}_0$. Therefore, supposing that $\mathcal{NF}_0$ is consistent, $\mathcal{NF}_1^*$ cannot be trivial. For instance, $\emptyset \in \emptyset$ is not a theorem of $\mathcal{NF}_1^*$, since $f(\emptyset \in \emptyset) = \emptyset \in \emptyset$ is not provable in $\mathcal{NF}_0$. ■

THEOREM 138. *If $\mathcal{NF}_0$ is consistent, then $\mathcal{NF}_1$ is nontrivial.*

Proof. The consistency of $\mathcal{NF}_0$ implies the nontriviality of $\mathcal{NF}_1$ because this system is weaker than $\mathcal{NF}_1^*$. ■

THEOREM 139. *If $\mathcal{NF}_0$ is consistent, then all the inconsistent systems $\mathcal{NF}_n$, $1 \leq n < \omega$, are nontrivial.*

Proof. It suffices to note that the theories $\mathcal{NF}_n$, $1 \leq n < \omega$, are weaker than $\mathcal{NF}_1^*$. ■

By slightly changing the proof of Theorem 138, we can prove the following result: if $\mathcal{NF}_0$ is consistent, then the system obtained from $\mathcal{NF}_n$, $1 \leq n < \omega$, by adding axioms guaranteeing the existence of the sets of all non-k-circular sets ($k = 1, 2, \ldots$) is nontrivial. For example, the set of all non-3-circular sets is the following set: $\{x : \neg\exists y_1 \exists y_2 \exists y_3 (x \in y_1 \wedge y_1 \in y_2 \wedge y_2 \in y_3 \wedge y_3 \in x)\}$.

$\mathcal{NF}_\omega$ is like $\mathcal{NF}_1$, except that its underlying logic is $C_\omega^=$ instead of $C_1^=$.

$\mathcal{NF}_\omega$ is weaker than $\mathcal{NF}_1$, so if $\mathcal{NF}_0$ is consistent, it is nontrivial.

3.2 Zermelo-Fraenkel-like systems

Starting with the Zermelo-Fraenkel system, we can introduce a new hierarchy of paraconsistent set theories $\mathcal{ZF}_n$, $1 \leq n \leq \omega$, similar to the hierarchy $\mathcal{NF}_n$, $1 \leq n \leq \omega$. Instead of ZF, we could have employed any other classical system of set theory, or type theory, as the first system in the hierarchy (see [da Costa and Alcântara, 1982] and [da Costa and Alcântara, 1986]). Among the several versions of ZF that can be used, perhaps the best form is Church's [Church, 1974], which admits the existence of the universal set.

Let us call the latter theory $\mathcal{ZF}_0$. The language L of $\mathcal{ZF}_0$ is that of $C_1^=$, but with only one specific (binary) predicate symbol: $\in$ (membership). The syntactic notions of L are obvious adaptations of those of $C_0^=$. In addition, $\notin$ has its standard meaning, and the description symbol ι is introduced by contextual definition, following Russell. In order to state the postulates of $\mathcal{ZF}_0$, we need some definitions.

DEFINITION 140.

1. $\{x : \alpha(x)\} =_{\text{def}} \iota y \forall x(x \in y \leftrightarrow \alpha(x))$

2. $\emptyset =_{\text{def}} \{x : x \notin x\}$

3. $\mathcal{V} =_{\text{def}} \{x : x = x\}$

4. $x \subset y \leftrightarrow \forall z(z \in x \rightarrow z \in y)$

5. $\{x\} =_{\text{def}} \{y : y = x\}$

6. $\mathcal{P}(x) =_{\text{def}} \{y : y \subset x\}$

7. $\bar{x} =_{\text{def}} \{y : y \notin x\}$

8. $\exists\{x : \alpha(x)\} \leftrightarrow \exists y \forall x(x \in y \leftrightarrow \alpha(x))$

We easily define other basic notions such as $x \cup y$, $x \cap y$, $\{x, y\}$, $\langle x, y\rangle$, $\cup x$, $\cap x$, relation, function, etc.

DEFINITION 141.

1. $trans(x) \leftrightarrow \forall y(y \in x \rightarrow y \subset x)$ (transitive set)

2. $conn(x) \leftrightarrow \forall y \forall z(y \in x\ z \in x \rightarrow (z = y \vee y \in z \vee z \in y))$ (connected set)

3. $wf(x) \leftrightarrow (x \neq \emptyset \rightarrow \exists y(y \in x \wedge x \cap y = \emptyset))$ (well-founded set)

4. $ord(x) \leftrightarrow trans(x) \wedge conn(x) \wedge wf(x)$ (x is an ordinal)

5. $low(x) \leftrightarrow wf(x) \wedge$ x is equipotent to y (x is a low set)

The postulates of $\mathcal{ZF}_0$ are those of $C_0^=$ plus the following:

(P1) $\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y$ (Extensionality)

(P2) $\exists\{z : z = x \vee z = y\}$ (Pair Set)

(P3) $\exists\{z : \exists y(y \in x \wedge z \in y\}$ (Union Set)

(P4) $\exists \bigcap x$ (Intersection Set)

(P5) $\exists\{x :$ x is a finite ordinal$\}$ (Infinity)

(P6) Any formulation of the Axiom of Choice.

(P7) $low(z) \rightarrow \exists x\{x : \alpha(x) \wedge x \in z\}$ (Comprehension)

(P8) $\forall x \forall y \forall z((\alpha(x, y) \wedge \alpha(x, z) \rightarrow x = y) \wedge (\alpha(x, y) \wedge \alpha(z, y) \rightarrow x = z)) \wedge \forall y(y \in t \rightarrow \exists x \alpha(x, y)) \rightarrow (low(t) \rightarrow \exists v \forall x(x \in v \leftrightarrow \exists y(\alpha(x, y) \wedge y \in t)))$ (Replacement)

(P9) $low(x) \rightarrow \exists\{y : y \subset x\}$ (Power Set)

(P10) $\exists\{y : y \notin x\}$ (Complement)

$\mathcal{ZF}_0$ is really a strong set theory; the well-founded sets constitute a 'model' of usual Zermelo-Fraenkel set theory. Moreover, the universal set does exist in $\mathcal{ZF}_0$:

$$\vdash \exists\{x : x = x\},$$

that is, the set $\mathcal{V}$ such that $\vdash \forall x(x \in \mathcal{V})$ and, in particular, $\vdash \mathcal{V} \in \mathcal{V}$. The collection of all sets, plus $\emptyset$, $\mathcal{V}$, $\subset$, $\cup$, $\cap$ and form a complete Boolean algebra.

THEOREM 142. *In $\mathcal{ZF}_0$:*

a) $\vdash x \cup \overline{x} = \mathcal{V}$ *b)* $\vdash x \cap \overline{x} = \emptyset$ *c)* $\vdash \overline{\overline{x}} = x$ *d)* $\vdash \overline{\mathcal{V}} = \emptyset$

e) $\vdash \bigcap \mathcal{V} = \emptyset$ *f)* $\vdash \bigcup \mathcal{V} = \mathcal{V}$ *g)* $\vdash \bigcap \emptyset = \mathcal{V}$ *h)* $\vdash \bigcup \emptyset = \emptyset$

i) $\vdash \neg wf(\mathcal{V})$ *j)* $\vdash \neg low(\mathcal{V})$ *k)* $\vdash \overline{\emptyset} = \mathcal{V}$

l) $x \subset y \to \overline{y} \subset \overline{x}$ *m)* $\vdash wf(\{\mathcal{V}\})$ *n)* $\vdash wf(\emptyset)$ *o)* $\vdash low(\emptyset)$

p) $\vdash$ *f is a function* $\wedge$ *$low(dom(f)) \to low(range(f))$, where $dom(f)$ and $range(f)$ stand for the domain and the range of f respectively.*

Let us call α^{wf} the formula obtained from the formula α by restricting its variables by the condition $wf(\ )$. Then we have the following result, whose proof is immediate:

THEOREM 143. *If α is a closed theorem of the standard ZF, then α^{wf} is a theorem of $\mathcal{ZF}_0$.*

As Church notes, it would be interesting to investigate the extension of $\mathcal{ZF}_0$ with the introduction of new postulates; for example, similar to the postulates of Quine's NF [Church, 1974]. A system of this type has already been studied in [da Costa, 1969].

Let us describe the system $\mathcal{ZF}_1$. This system is related to $\mathcal{ZF}_0$ as $C_1^=$ is related to $C_0^=$. (Recall that the latter is the standard first-order predicate calculus with identity.) So, we should have, among other things, that: (a) $\mathcal{ZF}_1$ should be partially included in $\mathcal{ZF}_0$, though the latter is also to be contained, in a certain sense, in the former. (b) $\mathcal{ZF}_1$ should be consistent, but can be used as the basis for inconsistent but nontrivial theories. In particular, in $\mathcal{ZF}_1$, we should be able to define 'inconsistent' set-theoretical structures, such as Russell set and Russell relations (see below), and other more complex structures (e.g. inconsistent arithmetics, and 'inconsistent' groups, among others).

$\mathcal{ZF}_1$ is constructed as follows: its language, called L, and the logical postulates are those of $C_1^=$. The basic set-theoretical concepts are analogous to those of $\mathcal{ZF}_0$, although the concepts involving negation give raise to two notions: one involving the weak negation ($\neg$), the other involving the strong negation ($\neg^\star$). In general, the symbols for those negations will differ only by the fact that the strong versions are starred. For instance, we have two empty sets: $\emptyset =_{\text{def}} \{x : \neg(x = x)\}$, and $\emptyset^\star =_{\text{def}} \{x : \neg^\star(x = x)\}$.

Each specific axiom or axiom scheme of $\mathcal{ZF}_0$ yields two corresponding axioms, or axiom schemas, of $\mathcal{ZF}_1$: one with the strong negation and another with the weak one. (We suppose that negation does occur essentially; otherwise, the postulate of $\mathcal{ZF}_0$ yields only one of $\mathcal{ZF}_1$, having the same syntactic form.)

It is important to note that, for example, $\mathcal{NF}_1$ is inconsistent, while $\mathcal{ZF}_1$ is apparently consistent, but nontrivial. However, it can be used as the basis for inconsistent but nontrivial theories.

The following results can be proved without difficulty. We emphasize, once again, that $\mathcal{ZF}_1$ is part of $\mathcal{ZF}_0$; but $\mathcal{ZF}_0$ is also contained, in certain sense, in $\mathcal{ZF}_1$.

THEOREM 144.

1. *Let* α *be a sentence of L and* $\alpha^\star$ *the sentence obtained from* α *by replacing* $\neg$ *by* $\neg^\star$*. Then* α *is a theorem of* $\mathcal{ZF}_0$ *iff* $\alpha^\star$ *is a theorem of* $\mathcal{ZF}_1$.

2. $\mathcal{ZF}_1$ *is consistent (with regard to* $\neg$ *or* $\neg^\star$*) iff* $\mathcal{ZF}_0$ *is consistent.*

3. *If* α *is a closed theorem of* $\mathcal{ZF}_0$*, then* $\forall x \forall y((x \in y)^o \wedge (x = y)^o) \rightarrow \alpha$ *is a theorem of* $\mathcal{ZF}_1$.

In the next subsection, we will show how $\mathcal{ZF}_1$ can be used as a paraconsistent basis for inconsistent but nontrivial theories (and, loosely speaking, for inconsistent mathematics).

As a final remark, note that in the set theories described above, we can construct semantics for C-logics, where the syntactic metalinguistic level can be considered as classical (in the sense that all the syntactic propositions are well-behaved; see [da Costa *et al.*, 1998a] for further references).

3.3 Russell sets and relations

By adapting the standard notion of a mathematical structure [Bourbaki, 1968], it is possible to define the concept of a paraconsistent structure [da Costa, 1989]. However, instead of developing here a general theory of paraconsistent structures, we will consider some particular cases, and show how a certain kind of paraconsistent mathematics can be developed.

DEFINITION 145 (Russell set). $R =_{\mathsf{def}} \{x : x \notin x\}$

The first structure to be studied is

(10) $\mathfrak{R} = \langle \mathcal{V}, R \rangle$,

characterized by the axiom

(11) $\exists R \wedge \forall x \exists \mathcal{P}(x)$,

where $\mathcal{V}$ is the domain (the universal set) of $\mathfrak{R}$, and R is its sole predicate; $\mathcal{P}(x)$ denotes the power set of x.

We emphasize that the following constructions are done in $\mathcal{ZF}_1$. Thus, we can make free use of the valid schemas and rules of $C_1^=$.

THEOREM 146. $\vdash R \in R \wedge R \notin R$

Proof. Given the definition of R, $x \in R \leftrightarrow x \notin x$. Hence, replacing x for R, we get $R \in R \leftrightarrow R \notin R$. However, if $R \in R$, it follows that $R \notin R$, and if $R \notin R$, we obtain $R \notin R$. Therefore, by excluded middle, $R \notin R$. Similarly, we prove that $R \in R$ by assuming that $R \notin R$. ■

THEOREM 147. $\vdash y \in \{x\} \leftrightarrow y = x$

Proof. In $\mathcal{ZF}_1$, we have that: $\exists\{x\}$. Therefore, $\vdash y \in \{x\} \leftrightarrow y = x$, by the definition of $\{x\}$. ■

THEOREM 148. $\vdash x \in R \rightarrow \{x\} \in R$

Proof. Either $\{x\} \notin \{x\}$ or $\{x\} \in \{x\}$. In the first case, $\{x\} \in R$, by the definition of R. In the second case, $\{x\} = x$, and given the hypothesis, $\{x\} \in R$. ■

THEOREM 149. $\vdash x, y \in R \rightarrow \{x, y\} \in R$

Proof. Either $\{x, y\} \notin \{x, y\}$ or $\{x, y\} \in \{x, y\}$. In the first case, $\{x, y\} \in R$, by the definition of R. In the second case, either $\{x, y\} = x$ or $\{x, y\} = y$, and given the hypothesis, $\{x, y\} \in R$. ■

THEOREM 150. $\vdash \{\{x, R\}\} \in R$

Proof. Either $\{\{x, R\}\} \in \{\{x, R\}\}$ or $\{\{x, R\}\} \notin \{\{x, R\}\}$. In the second case, it is immediate that $\{\{x, R\}\} \in R$. In the first case, by Theorem 147, it follows that $\{x, R\} = \{\{x, R\}\}$. Therefore, $x = R = \{x, R\}$, and given that $R \in R$, by Theorem 149, $\{x, R\} \in R$. Thus, by Theorem 148, $\{\{x, R\}\} \in R$. ■

THEOREM 151 (Arruda and Batens, 1982). $\bigcup R = \mathcal{V}$.

Proof. It suffices to prove that, for every x, $x \in \bigcup R$. Let us suppose that (i) $\{x, R\} \notin \{x, R\}$. Hence, $\{x, R\} \in R$ and, by the definition of the union set, $x \in \bigcup R$. On the other hand, if (ii) $\{x, R\} \in \{x, R\}$, then either $\{x, R\} = x$ or $\{x, R\} = R$. In the second case, it follows that $x \in \bigcup R$. If $\{x, R\} = x$, we have that $\{\{x, R\}\} = \{x\}$, and given that $\{\{x, R\}\} \in R$ (Theorem 150), it follows that $\{x\} \in R$. Hence, $x \in \bigcup R$. ■

This last theorem shows that a set theory with Russell set has, in general, a universal class.

THEOREM 152 (Arruda). $\ldots \subset \mathcal{P}(\mathcal{P}(R)) \subset \mathcal{P}(R) \subset R$.

Proof. If $x \in \mathcal{P}(R)$, then $x \subset R$. Now, either $x \notin x$ or $x \in x$. If $x \notin x$, then $x \in R$; if $x \in x$, given that $x \subset R$, it follows that $x \in R$. Therefore, $\mathcal{P}(R) \subset R$. Furthermore, if $x \in \mathcal{P}(\mathcal{P}(R))$, then $x \subset \mathcal{P}(R)$, and by the previous result, $x \subset R$; hence $x \in \mathcal{P}(R)$. Thus, $\mathcal{P}(\mathcal{P}(R)) \subset \mathcal{P}(R) \subset R$. It is now easy to complete the proof. ■

THEOREM 153. *In $\mathcal{ZF}_1$ plus the axiom (11):*

1. $\vdash \emptyset \in R, \{\emptyset\} \in R, \{\{\emptyset\}\} \in R \ldots$

2. $\vdash \exists x(x \notin R)$
3. $\vdash \mathcal{V} \notin R$
4. $\vdash x, y \in R \rightarrow \langle x, y\rangle \in R$
5. $\vdash \mathcal{V} = \mathcal{P}(\mathcal{V})$
6. $\vdash x \subset R \rightarrow x \in R$
7. $\vdash R \cup \overline{R} = \bigcup R$
8. $\vdash R \times R \subset R$
9. $\vdash \neg^{\star}\exists\{x : \neg^{\star}(x \in x)\}$

10 $\vdash \exists R \rightarrow \exists x(x \text{ is infinite})$

Given Theorem 150, it is possible to demonstrate that R is, as it were, an 'internal model' of the set theory in which we are working. Moreover, given that $\bigcup R = \mathcal{V}$, it follows that the existence of R implies the existence of infinite sets. The properties of R are by no means arbitrary. For example, it does not seem possible to prove that $(R \in R)^o$. Moreover, it is not possible to prove everything with regard to R without also proving, at the same time, that some classical, standard set theories are inconsistent (see [da Costa, 1964a] and [da Costa, 1986a]). Finally, $\mathcal{ZF}_1$ with the extra axiom $\exists R$ is nontrivial iff $\mathcal{ZF}_0$ is consistent.

Russell relations

DEFINITION 154 (Russell relations). Let $n = 1, 2, \ldots$, and $0 < i \leq n$. Then:

$$R_{n,i} =_{\text{def}} \{\langle x_1, \ldots, x_n\rangle : \langle x_1, \ldots, x_n\rangle \notin x_i\}.$$

By convention, we have: $\langle x\rangle = x$.

In particular, $R_{1,1} = R$. With $R_{n,i}$, assuming appropriate postulates, we can build structures similar to $\mathfrak{R}$ above. We can also prove the following results:

THEOREM 155.

1. $\vdash R_{n,i} \in R_{n,i} \wedge R_{n,i} \notin R_{n,i}$
2. $\vdash \mathcal{V} \times \cdots \times \mathcal{V} = \bigcup R_{n,i}$, where the product on the left has n terms. In particular, $\vdash \mathcal{V} \times \mathcal{V} \subset \cup \cup R_{2,1}$

THEOREM 156.

1. $\vdash \forall t(\langle \emptyset, t\rangle \in R_{2,1})$
2. $\vdash \cup \cup R_{2,1} = \mathcal{V}$

Proof. With regard to 1, by hypothesis, $\exists R_{2,1}$. Then, $\langle \emptyset, t\rangle \in R_{2,1} \leftrightarrow \langle \emptyset, t\rangle \notin \emptyset$. With regard to 2, since $\{\{\emptyset\}, \{\emptyset, t\}\} = \langle \emptyset, t\rangle \in R_{2,1}$, it follows that $\{\emptyset, t\} \in \cup R_{2,1}$ and $t \in \cup\cup R_{2,1}$. ■

Given the consistency of $\mathcal{NF}_0$ (see the previous subsection), if we add to $\mathcal{NF}_n$, $1 \leq n < \omega$, new postulates guaranteeing the existence of Russell relations, the resulting system is nontrivial.

3.4 Paraconsistent Boolean algebra

In various paraconsistent set theories, even in classical set theories such as Zermelo-Fraenkel, it is possible to consider intuitively a set as an ordered pair, in the classical sense, of sets that are part of a universal class $\mathcal{V}$. Thus, a set X is a pair $\langle X_1, X_2\rangle$, where:

(1) $x \in X$ iff $x \in X_1$

(2) $x \notin X$ iff $x \in X_2$

(3) $x \in X$ and $x \notin X$ is equivalent to $x \in X_1$ and $x \in X_2$

Given that the principle of excluded middle holds in certain paraconsistent set theories, it should be the case that $X_1 \cup X_2 = \mathcal{V}$. If $X_1 \cap X_2 = \emptyset$, a classical set is obtained. Let us consider then the collection of the sets just constructed on $\mathcal{V}$, which will be denoted by $\mathfrak{V}$. An element of $\mathfrak{V}$ is called a paraconsistent set, or a p-set. In what follows, we outline an algebra of p-sets $\mathfrak{V}$. We suppose that the p-sets are embedded in a classical set theory, for instance, ZF.

Suppose that $X = \langle X_1, X_2\rangle$ and $Y = \langle Y_1, Y_2\rangle$. In this case:

DEFINITION 157.

(i) $X \sqcup Y =_{\mathsf{def}} \langle X_1 \cup Y_1, X_2 \cap Y_2\rangle$

(ii) $X \sqcap Y =_{\mathsf{def}} \langle X_1 \cap Y_1, X_2 \cup Y_2\rangle$

(iii) $\mathbf{1} =_{\mathsf{def}} \langle \mathcal{V}, \emptyset\rangle$

(iv) $\mathbf{0} =_{\mathsf{def}} \langle \emptyset, \mathcal{V}\rangle$

(v) $\overline{X} =_{\mathsf{def}} \langle X_2, X_1\rangle$

(vi) $X \sqsubset Y =_{\mathsf{def}} X_1 \subset Y_1 \wedge Y_2 \subset X_2$

THEOREM 158.

$\vdash X \sqcup X = X$	$\vdash X \sqcap X = X$	$\vdash \overline{\overline{X}} = X$
$\vdash X \sqcup Y = Y \sqcup X$	$\vdash X \sqcap Y = Y \sqcap X$	$\overline{\boldsymbol{1}} = \boldsymbol{0}$
$\vdash (X \sqcup Y) \sqcup Z = X \sqcup (Y \sqcup Z)$	$\vdash (X \sqcap Y) \sqcap Z = X \sqcap (Y \sqcap Z)$	$\vdash \overline{\boldsymbol{0}} = \boldsymbol{1}$
$\vdash \boldsymbol{1} \sqcup X = \boldsymbol{1}$	$\vdash \boldsymbol{1} \sqcap X = X$	$\vdash \boldsymbol{0} \sqcap X = \boldsymbol{0}$
$\vdash \boldsymbol{0} \sqcup X = X$	$\vdash X \sqcup \overline{X} \sqsubset \boldsymbol{1}$	$\vdash \boldsymbol{0} \sqsubset X \sqcap \overline{X}$
$\vdash X \sqsubset X$	$\vdash \boldsymbol{0} \sqsubset X$	$\vdash X \sqsubset \boldsymbol{1}$
$\vdash X \sqsubset Y \wedge Y \sqsubset X \rightarrow X = Y$	$\vdash X \sqsubset Y \wedge Y \sqsubset Z \rightarrow X \sqsubset Z$	
$\vdash X \sqsubset X \sqcup \overline{X}$	$\vdash X \sqcap \overline{X} \sqsubset X$	

DEFINITION 159. The structure $\mathfrak{P} = \langle \mathfrak{V}, \sqcup, \sqcap, -, \mathbf{1}, \mathbf{0}\rangle$ is called a paraconsistent Boolean algebra.

Using this structure, it is possible to formalize several paraconsistent patterns of reasoning, just as with classical Boolean algebras, we can put in algebraic terms various classical inferences. We can also verify that the paraconsistent logical mechanism considered here does not exclude classical logic, but extends it in a certain sense — though under another point of view, it can be embedded into classical logical structures. Of course, this point is valid for particular categories of paraconsistent structures. It corroborates the fact that both paraconsistent logic and paraconsistent mathematics do not destroy either traditional logic or standard mathematics. Rather, the paraconsistent approach complements the classical developments of logic and mathematics, and in certain cases, extends them.

The structure of the paraconsistent Boolean algebra clearly is richer than the classical one. Thus, for instance, one can introduce two operators, π_1 and π_2, such that, given a p-set X, $\pi_1(X) = X_1$ and $\pi_2(X) = X_2$, where X_1 and X_2 are in another Boolean algebra, the classical algebra of the subsets of $\mathcal{V}$ etc.

When the structure $\mathfrak{P}$, in the above definition, is such that, for every $X = \langle X_1, X_2 \rangle$, it is the case that $X_1 \cap X_2 = \emptyset$, one obtains a Boolean algebra that essentially is the usual algebra of the subsets of $\mathcal{V}$.

In this way, it is possible to construct a general theory of paraconsistent structures (algebraic, topological, of order etc.), obtaining in this way a generalization of the traditional theory of structures, such as Bourbaki's. Moreover, as we will see below, paraconsistent structures, such as those described in this section, have been applied to several areas, such as computer science, artificial intelligence and logic programming (see, e.g., [Subrahmanian, 1987], [Blair and Subrahmanian, 1989a], [Blair and Subrahmanian, 1988b], and [Kifer and Subrahmanian, 1989]). [13] This provides a significant motivation for their study.

3.5 Paraconsistent mathematics

In this subsection, we outline a paraconsistent formulation of elementary differential and integral calculus, in which what we will call l'Hospital principle is true:

> 'Two finite quantities that differ by an infinitely small quantity are equal.' [14]

The paraconsistent nature of this principle is clear, even though we are not trying to advance an exegesis of l'Hospital's work (see [l'Hospital, 1996] and [Robinson, 1966]). To begin with, we describe two (classical) algebraic structures $\mathcal{A}$ and $\mathcal{A}^\star$.

The ring $\mathcal{A}$ The ring $\mathcal{A}$ is described as follows. Let $\mathbb{R}$ denote the field of real numbers and a an element of a fixed open interval I ($I \subset \mathbb{R}$). An *infinitesimal variable* is a real valued function defined on I that has limit zero in a (we can use right or left limits). The expression 'infinitesimal variable' is here employed inspired by the terminology of the classical French treatises, such as those by Picard and Gousart, as well as by Cauchy. The elements of $\mathcal{A}$ are ordered pairs $\langle r, f \rangle$, where $r \in \mathbb{R}$ and f is an infinitesimal variable.

---

[13] It is worth noting that paraconsistent structures have also been employed in quantum mechanics by Dalla Chiara and Giuntini (see [Dalla Chiara and Giuntini, 1989], [Dalla Chiara and Giuntini, 2000], and [Dalla Chiara and Giuntini, 2001]).

[14] In this context, it is worth also giving a quotation attributed to Johann Bernoulli (1667-1748), according to which 'a quantity that is increased or decreased by an infinitely small quantity is neither increased nor decreased'.

If $\langle r, f\rangle$ and $\langle s, g\rangle$ belong to $\mathcal{A}$, then $\langle r, f\rangle = \langle s, g\rangle$ if $r = s$ and $f = g$. Addition is defined as follows: $\langle r, f\rangle + \langle s, g\rangle =_{\text{def}} \langle r + s, f + g\rangle$. By an abuse of language, we take $0 = \langle 0, 0\rangle$, where the second zero in $\langle 0, 0\rangle$ is the identically zero function in I. We also take $-\langle r, f\rangle =_{\text{def}} \langle -r, -f\rangle$, and $\langle r, f\rangle - \langle s, g\rangle =_{\text{def}} \langle r, f\rangle + (-\langle s, g\rangle) = \langle r - s, f - g\rangle$.

The following additional definitions are needed: $\langle r, f\rangle \cdot \langle s, g\rangle =_{\text{def}} \langle rs, rg + fs + fg\rangle$ and $1 =_{\text{def}} \langle 1, 0\rangle$. Furthermore, if $r \neq 0$, we put $\langle r, f\rangle^{-1} =_{\text{def}} \langle r^{-1}, -f/r(f + r)\rangle$ and $\langle s, g\rangle \div \langle r, f\rangle =_{\text{def}} \langle s, g\rangle \cdot \langle r, f\rangle^{-1}$.

The elements of $\mathcal{A}$ with the operation + constitute a commutative group, and with ×, a commutative semi-group with unity. Moreover, multiplication is distributive in relation to addition. Therefore, $\mathcal{A}$ is a commutative ring with unity. The elements of $\mathcal{A}$ are called *hyper-reals*. If we identify $\langle r, 0\rangle$ with r, then the field $\mathbb{R}$ is contained in $\mathcal{A}$ as a sub-ring. We write $\langle r, f\rangle$ as $r + \langle 0, f\rangle$. A pair such as $\langle 0, f\rangle$ is called an *infinitesimal*, and will be usually denoted by small Greek letters. Any element of $\mathcal{A}$, $\langle r, f\rangle$, then, is the sum of a standard real and an infinitesimal: $\langle r, f\rangle = r + \varepsilon$, where $\varepsilon = \langle 0, f\rangle$.

Division can be extended to infinitesimals. In fact, let us suppose that $\kappa = \langle 0, f\rangle$, $\lambda = \langle 0, g\rangle$, and $\lim f/g = r$ in $\mathcal{A}$. Also, let $\kappa/\lambda = \langle r, f/g - r\rangle$, where the variable g is supposed not to assume the value 0. Therefore, making $h = f/g - r$, it is easy to check that the infinitesimal $\varepsilon = \langle 0, h\rangle$ is such that $\kappa = \lambda \cdot \varepsilon$. When $\lim f/g = \infty$, or does not exist, the quotient κ/λ is not defined.

We introduce in $\mathcal{A}$ a relation of inequality, <, so that $\langle r, f\rangle < \langle s, g\rangle$ if $r < s$, or in case $r = s$, if $f < g$ (the variable f is less than the variable g in all points of I).

The set of all hyper-reals of the form $\langle r, f\rangle$ is called *the monad* of $r \in \mathbb{R}$. The *order* of an infinitesimal ε in relation to another infinitesimal κ is defined without difficulty. Given a function $f : \mathbb{R} \mapsto \mathbb{R}$, it can be extended so that $dom(f) = \mathcal{A}$, under the hypothesis that $\lim_{x\to t} f(x)$ exists and is finite for any $t \in \mathbb{R}$. We take $f(r + \varepsilon) = b + \delta$, where $b = \lim_{x\to r} f(x)$ and δ is an infinitesimal obviously defined.

The basic concepts of the differential calculus can then be defined in terms of infinitesimals. For example, the expression:

$$\lim_{x\to r} f(x) = b$$

means that $f(r + \varepsilon) = b + \delta$, for every infinitesimal ε (δ is also an infinitesimal), where $r \in \mathbb{R}$. Similarly, the derivative of f is defined as follows:

$$f(r + \varepsilon) - f(r) = f'(r) \cdot \varepsilon + \delta,$$

for any infinitesimal ε, δ being an infinitesimal of order higher than that of ε. The properties of limits, derivatives etc. are easily proved from the properties of infinitesimals.

The quasi-ring $\mathcal{A}^\star$ The quasi-ring $\mathcal{A}^\star$ is obtained from $\mathcal{A}$ by the introduction of infinite 'numbers'. An *infinite variable* is a standard real-variable function, defined in I, and divergent at $a \in I$.

The pair $\langle v, 0\rangle$ is called an infinite *hyper-real* number when v is an infinite variable. The (finite) elements of $\mathcal{A}$ and the infinite numbers form $\mathcal{A}^\star$. Two infinite numbers $\langle v, 0\rangle$ and $\langle u, 0\rangle$ are equal if $v = u$. The operations of $\mathcal{A}^\star$ are defined as in $\mathcal{A}$, but extended as follows:

Addition

1. If $\langle v, 0\rangle$ is infinite and $\langle k, f\rangle$ is finite, then $\langle v, 0\rangle + \langle k, f\rangle = \langle k, f\rangle + \langle v, 0\rangle = \langle v + k, 0\rangle$.

2. If $\langle v, 0\rangle$ and $\langle u, 0\rangle$ are both infinite, then $\langle v, 0\rangle+\langle u, 0\rangle = \langle u, 0\rangle+\langle v, 0\rangle = \langle v+u, 0\rangle$, whenever $\lim(v+u) = \infty$; $\langle v, 0\rangle+\langle u, 0\rangle = \langle u, 0\rangle+\langle v, 0\rangle = \langle k, f\rangle$ when $\lim(v+u) = k$ (standard real), where $f = k - (u + v)$. Otherwise, $\langle v, 0\rangle + \langle u, 0\rangle$ and $\langle u, 0\rangle + \langle v, 0\rangle$ are not defined.

Subtraction By definition, the opposite of an infinite $\langle v, 0\rangle$ is $\langle -v, 0\rangle$. Then, the difference of two elements of $\mathcal{A}^\star$ is defined as the sum of the first with the opposite of the second.

Multiplication

1. If $\langle v, 0\rangle$ and $\langle u, 0\rangle$ are infinite, then $\langle v, 0\rangle \cdot \langle u, 0\rangle = \langle u, 0\rangle \cdot \langle v, 0\rangle = \langle uv, 0\rangle$.

2. If $\langle v, 0\rangle$ is infinite and $\langle k, f\rangle$ is finite, but not an infinitesimal, then $\langle v, 0\rangle\cdot\langle k, f\rangle = \langle k, f\rangle \cdot \langle v, 0\rangle = \langle vk, 0\rangle$

3. If $\langle v, 0\rangle$ is infinite and $\langle 0, f\rangle$ is an infinitesimal, then $\langle v, 0\rangle \cdot \langle 0, f\rangle = \langle 0, f\rangle \cdot \langle v, 0\rangle = \langle vf, 0\rangle$, when $\lim(vf) = \infty$; if $\lim(vf) = 0$, the product is equal to $\langle 0, vf\rangle$. Otherwise, the product is not defined.

Division Let $\langle v, 0\rangle$ be an infinite number satisfying obvious conditions; $\langle v, 0\rangle^{-1}$ is the infinitesimal $\langle 0, v^{-1}\rangle$. If $\langle 0, f\rangle$ is an infinitesimal satisfying appropriate conditions, then $\langle 0, f\rangle^{-1}$ is the infinite $\langle f^{-1}, 0\rangle$. The quotient of two elements of $\mathcal{A}^\star$ is the product of the first by the inverse of the second.

The relation $<$ can be extended to $\mathcal{A}^\star$ without difficulty. (However, $\mathcal{A}$ and $\mathcal{A}^\star$ are not Archimedean structures.) In terms of infinitesimals and infinities, we can express in $\mathcal{A}^\star$ the basic ideas and results of the infinitesimal calculus (for analogous views, see [Schmieden and Laugwitz, 1958], [Laugwitz, 1961a], and [Laugwitz, 1961b]). In particular, we can get something like de l'Hospital's principle on curves, according to which a smooth curve may be analyzed into an infinite number of infinitesimal straight lines ([l'Hospital, 1996], and [Robinson, 1966]).

The classical theory of infinitesimals and infinite quantities, as well as Du Bois Reymond's theory of orders of magnitude concerning the asymptotic behavior of functions (see [Borel, 1950, Note II]), are translatable in the language of $\mathcal{A}^\star$. The same is true in connection with other topics of pure and applied mathematics, such as, the theory of differential equations and Fourier series. Loosely speaking, $\mathcal{A}^\star$ is a model of a theory with infinitesimals and infinite quantities. It is also worth noting that natural suppositions about the orders of infinitesimals and infinities give raise to propositions that are undecidable in Zermelo-Fraenkel set theory (see [Ellentuck, 1975]).

To describe a paraconsistent model for the differential and integral calculus, we start with the language L in which we can treat the central notions of this calculus. L is essentially Manin's $\mathrm{L_2Real}$ [Manin, 1977, p. 109], conveniently extended by the introduction of names, in the sense of Shoenfield (see [Shoenfield, 1967]), for all elements of $\mathcal{A}$ (or of $\mathcal{A}^\star$). So, L is composed of the following primitive symbols: (1) individual variables; (2) variables for functions of one variable; (3) individual constants: the names of the elements of $\mathcal{A}$ (or of $\mathcal{A}^\star$); (4) the symbols $+$ and $\times$ for binary operations; (5) two binary relation symbols: $\equiv$ and $<$; (6) the connectives: $\rightarrow$, $\wedge$, $\vee$, $\leftrightarrow$ and $\neg$; and (7) the quantifiers: $\forall$ and $\exists$; (8) parentheses.

Terms of L: (a) All individual variables, individual constants, and function variables are terms. (b) If F is a function variable and t is a term, then $F(t)$ is a term. (c) If t_1 and t_2 are terms, so are $t_1 + t_2$ and $t_1 \times t_2$.

Formulas of L: If t_1 and t_2 are terms, then $t_1 \equiv t_2$ and $t_1 < t_2$ are atomic formulas. The remaining formulas are defined as usual, but quantification over function variables is allowed. The common syntactic notions such as those of bound and free variables, sentence etc. are defined in the usual way.

The following clauses define when v is a paraconsistent valuation of L:

1. Names denote the corresponding elements of $\mathcal{A}$ (or of $\mathcal{A}^\star$).

2. Let $t_1 < t_2$ be an atomic sentence. In this case, $v(t_1 < t_2) = 1$ if $t_1 < t_2$ is true in $\mathcal{A}$ (or in $\mathcal{A}^\star$). Otherwise, $v(t_1 < t_2) = 0$.

3. $v(t_1 \equiv t_2) = 1$ iff $t_1 - t_2$ is infinitesimal with respect to ε (ε is an infinitesimal).

4. Let $t_1 \equiv t_2$ be an atomic sentence. In this case, $v(t_1 \not\equiv t_2) = v(\neg(t_1 \equiv t_2)) = 1$ if $t_1 \neq t_2$ in $\mathcal{A}$ (or in $\mathcal{A}^\star$); otherwise, $v(t_1 \not\equiv t_2) = 0$.

5. Similarly, we define the value of v for any sentence of L, replacing $\equiv$ and $\not\equiv$ in the sentence by convenient symbolic combinations, respectively as in 3 and 4.

If a sentence F is undecidable in the usual set theory plus the axioms of $\mathcal{A}$ (or of $\mathcal{A}^\star$), the value of F is chosen arbitrarily, in such a way that v be the characteristic function of a maximal nontrivial set of sentences. Therefore, v is really a valuation in L.

It is immediate to verify that we may have, for some terms t_1 and t_2, that $v(t_1 \equiv t_2) = v(t_1 \not\equiv t_2) = 1$. In fact, v constitutes a paraconsistent valuation and determines a *model* of the infinitesimal calculus in which de l'Hospital's principle about finite quantities that differ by an infinitely small increment holds. Moreover, in the case of $\mathcal{A}^\star$, the model is such that even de l'Hospital's second principle, on smooth curves, happens to be valid (when conveniently interpreted).

The method sketched here to obtain v is analogous to those of Chris Mortensen (see [Mortensen, 1995]), and of synthetic differential geometry (see [Bell, 1988] and [Bell, 1995]). For the treatment of functions of several variables, we have to strengthen L with the introduction of functional variables (see [Scott, 1967]).

In summary, we may say that v is a paraconsistent structure, which we will call the *de l'Hospital structure*. On the paraconsistent version of the infinitesimal calculus outlined here, see also [da Costa *et al.*, 1998a]. Further developments could start with the paradoxes of Burali-Forti and of Cantor, which naturally motivate the definition of interesting paraconsistent structures. The development of paraconsistent arithmetics, in turn, offers no difficulty at all.

4 JAŚKOWSKI'S LOGIC

4.1 Jaśkowski's discussive logic

As noted above, following a suggestion by J. Łukasiewicz, Stanislaw Jaśkowski (1906-1965) was the first logician to construct a system of paraconsistent propositional calculus (see [Jaśkowski, 1948], [Jaśkowski, 1949], [Jaśkowski, 1969], and [da Costa and Doria, 1995]). Jaśkowski motivated his *discussive* logic (sometimes also referred to as *discursive* logic) by the need for addressing three issues: (i) to systematize theories that contain contradictions, such as dialectics; (ii) to study theories where there are contradictions originated by vagueness, and (iii) to investigate empirical theories whose postulates or basic assumptions are contradictory (see [Arruda, 1980] and [Arruda, 1989]).

Later, da Costa and Dubikajtis extended the discussive propositional calculus to first- and higher-order predicate calculi (see [da Costa and Dubikajtis, 1968] and [da Costa and Dubikajtis, 1977]; see also [Kotas, 1971], [Kotas, 1974], and [Kotas, 1975]). Recently, discussive logic has been applied to the theory of pragmatic truth (as we will discuss below), to the foundations of physics (see [da Costa and Doria, 1995] and also below), and in the philosophy of science (for a general account, see [da Costa and French, 2003]). In this section, we will sketch the main ideas related to Jaśkowski's discussive logic.

Let us call $\mathcal{J}$ the discussive propositional calculus whose language $\mathcal{L}$ and notations are those one the modal system S5. We pose: $\Diamond\Gamma =_{\mathsf{def}} \{\Diamond\alpha : \alpha \in \Gamma\}$. So, $\mathcal{J}$ can be semantically defined as

$$\Gamma \models_{\mathcal{J}} \alpha \text{ iff } \Diamond\Gamma \models_{S5} \Diamond\alpha,$$

where the notation has an obvious meaning. It is immediate that

THEOREM 160.

(1) $\models_{\mathcal{J}} \alpha$ *iff* $\models_{S5} \Diamond\alpha$

(2) $\Gamma \models_{\mathcal{J}} \alpha$ *iff there are* $\gamma_1, \ldots, \gamma_n$ in Γ *such that* $\models_{S5} \Diamond\gamma_1 \wedge \ldots \wedge \Diamond\gamma_n \to \Diamond\alpha$

COROLLARY 161. $\Gamma \models_{\mathcal{J}} \alpha$ *iff there is a finite set* $\{\gamma_1, \ldots, \gamma_n\} \subseteq \Gamma$ *such that* $\{\gamma_1, \ldots, \gamma_n\} \models_{\mathcal{J}} \alpha$.

COROLLARY 162. *If* $\models_{S5} \alpha$, *then* $\models_{\mathcal{J}} \alpha$.

Proof. It is enough to note that if $\models_{S5} \alpha$, then $\models_{\mathcal{J}} \Diamond\alpha$. ■

Given the definitions and results above, we can see that $\mathcal{J}$ accomplishes Jaśkowski's main intensions.

$\mathcal{J}$ has several axiomatizations (see [da Costa, 1975] and [da Costa and Dubikajtis, 1977]); the one presented here was introduced in [da Costa and Doria, 1995]. The postulates are:

(J1) If α is an axiom of S5, then $\Box\alpha$.

(J2) $\Box\alpha, \Box(\alpha \to \beta) / \Box\beta$

(J3) $\Box\alpha / \alpha$

(J4) $\Diamond\alpha / \alpha$

(J5) $\Box\alpha / \Box\Box\alpha$

LEMMA 163. *If $\vdash_{\mathcal{J}} \alpha$ means that α is provable in $\mathcal{J}$, then: if $\vdash_{\mathcal{J}} \alpha$, then $\models_{S5} \Diamond\alpha$.*

Proof. By induction on the length of a given derivation of α in $\mathcal{J}$. ■

LEMMA 164. *If $\models_{S5} \alpha$, then $\vdash_{\mathcal{J}} \Box\alpha$.*

Proof. By induction on the length of a given derivation of α in S5. ■

THEOREM 165. *$\vdash_{\mathcal{J}} \alpha$ iff $\models_{\mathcal{J}} \alpha$.*

Proof. If $\vdash_{\mathcal{J}} \alpha$, then by Lemma 163, $\models_{S5} \Diamond\alpha$. So, by definition, $\models_{\mathcal{J}} \alpha$. Conversely, if $\models_{\mathcal{J}} \alpha$, then by definition $\models_{S5} \Diamond\alpha$. So, by Lemma 164, $\vdash_{\mathcal{J}} \Box\Diamond\alpha$. By postulate (J3), $\vdash_{\mathcal{J}} \Diamond\alpha$, and by postulate (J4), $\vdash_{\mathcal{J}} \alpha$. ■

DEFINITION 166. We write $\Gamma \vdash_{\mathcal{J}} \alpha$ iff there are $\gamma_1, \ldots, \gamma_n$ such that $\vdash_{\mathcal{J}} \Diamond\gamma_1 \wedge \ldots \wedge \Diamond\gamma_n \to \Diamond\alpha$.

THEOREM 167. *$\Gamma \vdash_{\mathcal{J}} \alpha$ iff $\Gamma \models_{\mathcal{J}} \alpha$.*

Proof. Immediate consequence of the definition above and Theorem 165. ■

THEOREM 168. *Modus Ponens, that is, the rule $\alpha, \alpha \to \beta / \beta$, is not valid in $\mathcal{J}$.*

Proof. The uniform predicate calculus $\mathcal{U}$ is a subcalculus of the monadic calculus; it is, the first-order predicate calculus dealing only with unary predicates in which there is only one individual variable, say x (see [Church, 1956]). There is an obvious bijection between the set of formulas of $\mathcal{U}$ and the language of S5: given a formula α of $\mathcal{U}$, we obtain the corresponding formula α' of the language of S5 by replacing any subformula $P_i(x)$ of α by p_i, and any universal quantification $\forall x$ by $\Box$. We can show that if α is a formula of $\mathcal{U}$ and α' is its corresponding formula in S5, then $\models_{\mathcal{U}} \alpha$ iff $\models_{S5} \alpha'$ (see [da Costa and Doria, 1995]). To prove the theorem, it is enough to note that $\exists x\alpha(x)$ and $\exists x(\alpha(x) \to \beta(x))$ do not imply that $\exists x\beta(x)$ in $\mathcal{U}$. ■

THEOREM 169. *(a) The rules $\alpha,\beta / \alpha \wedge \beta$ and $\alpha, \neg\alpha / \beta$ are not valid in $\mathcal{J}$. (b) The deduction theorem $\Gamma, \alpha \vdash_{\mathcal{J}} \beta \Rightarrow \Gamma \vdash_{\mathcal{J}} \alpha \to \beta$ is not true in $\mathcal{J}$.*

These results show that $\mathcal{J}$ can be used to accommodate inconsistent sets of premisses while avoiding triviality; that is, it is paraconsistent.

DEFINITION 170.

(1) [Discussive Implication] $\alpha \to_d \beta =_{\text{def}} \Diamond\alpha \to \beta$

(2) [Discussive Conjunction] $\alpha \wedge_d \beta =_{\text{def}} \Diamond\alpha \wedge \beta$

(3) [Discussive Equivalence] $\alpha \leftrightarrow_d \beta =_{\text{def}} (\alpha \to_d \beta) \wedge_d (\beta \to_d \alpha)$

(4) [Impossibility] $\nabla\alpha =_{\text{def}} \neg\Diamond\alpha$

THEOREM 171. *The connectives* $\to_d$, $\vee$, $\wedge_d$, $\leftrightarrow_d$ *and* $\leftrightarrow_d$ *have all the classic properties of* $\to$, $\vee$, $\wedge$, $\leftrightarrow$ *and* $\neg$ *respectively.*

This shows that the classical propositional calculus is naturally embedded in $\mathcal{J}$. However, as the next theorem shows, several rules and formulas do not hold in this calculus.

THEOREM 172. *The following formulas and rules are not valid in* $\mathcal{J}$.

(1) $\alpha \to_d (\beta \to_d \alpha \wedge \beta)$

(2) $\alpha \to_d (\neg\alpha \to_d \beta)$

(3) $(\alpha \wedge \beta \to_d \gamma) \to_d (\alpha \to_d (\beta \to_d \gamma))$

(4) $\Gamma, \alpha \models_{\mathcal{J}} \beta$ and $\Gamma, \alpha \models_{\mathcal{J}} \neg\beta \Rightarrow \Gamma \models_{\mathcal{J}} \neg\alpha$

(5) $(\alpha \leftrightarrow_d \neg\alpha) \to_d \beta$

(6) $(\alpha \to_d \neg\alpha) \to_d \neg\beta$

THEOREM 173.

(1) $\mathcal{J}$ is decidable.

(2) $\mathcal{J}$ has no finite characteristic matrix, but has the finite model property.

Proof. With regard to (1), the decidability of $\mathcal{J}$ follows from S5's decidability. In fact, if α is a formula of $\mathcal{J}$, use Definition 170 to obtain a formula β of S5 by eliminating the discussive connectives. Now, since β is a theorem of $\mathcal{J}$ iff $\Diamond\beta$ is a theorem of S5, use a decision procedure for S5 to verify whether $\Diamond\beta$ is a theorem of S5. With regard to (2), it follows from the fact that S5 has the finite model property, even though it has no finite characteristic matrix (for details, see [da Costa, 1975]). ■

$\mathcal{J}$ has (at least) two equivalent semantics. One based on Kripke structures, the other based on the notion of Hanle algebra. A Hanle algebra is a structure $\mathcal{H} = \langle A, -, \wedge, \star \rangle$, where $\langle A, -, \wedge \rangle$ is a Boolean algebra and $\star$ is an unary operator over A such that: $\star 1 = 1$ and $\star x = 0$, for all $x \neq 0$, where 0 and 1 are, respectively, the first and the last elements of the Boolean algebra. We can prove that the following sentences are equivalent (where α is a formula of S5): (i) α is a theorem of S5; (ii) α is valid in every Hanle algebra, and (iii) α is valid in every finite Hanle algebra (see [da Costa and Doria, 1995]). From these statements, it is easy to derive a semantics for $\mathcal{J}$. Nevertheless, in what follows, we will make reference only to the semantics based on Kripke structures.

A Kripke semantics for $\mathcal{J}$ can be constructed by the same steps we take to build a model theory for S5, and the corresponding results can be proved.

DEFINITION 174. Let Γ be a set of formulas of the language of $\mathcal{J}$. Then:

(1) $\overline{\Gamma} =_{\mathsf{def}} \{\alpha : \Gamma \vdash_{\mathcal{J}} \alpha\}$

(2) *If* $\overline{\Gamma}$ *is the set of all formulas, then* Γ *is trivial; otherwise, it is nontrivial.*

(3) *If there is a formula* α *such that* $\Gamma \vdash_{\mathcal{J}} \alpha$ *and* $\Gamma \vdash_{\mathcal{J}} \neg\alpha$, *then* Γ *is inconsistent; otherwise, it is consistent.*

(4) *If there is a formula* α *such that* $\Gamma \vdash_{\mathcal{J}} \alpha$ *and* $\Gamma \vdash_{\mathcal{J}} \nabla\alpha$, *then* Γ *is strongly inconsistent.*

THEOREM 175. *There are inconsistent, but nontrivial sets of formulas.*

Proof. If p is a propositional variable, then $\{p, \neg p\}$ is inconsistent, but nontrivial. ■

A Kripke structure is an ordered pair $K = \langle W, \upsilon\rangle$, where W is a non-empty set whose elements are called *worlds*, and υ is a mapping $\upsilon : W \times P \mapsto \{0, 1\}$, where P is the set of propositional variables of the language of $\mathcal{J}$. As usual, if $\upsilon(w, p) = 1 (= 0)$, we say that p is *true* (*false*) in the world w.

For brevity's sake, let us take $\vee$, $\neg$, and $\Box$ as the primitive connectives of our language. If α is a formula, then we say that K, w *force* α, and write $K, w \Vdash \alpha$, according to the following recursive definition:

(i) $K, w \Vdash p$ iff $\upsilon(w, p) = 1$

(ii) $K, w \Vdash \beta \vee \gamma$ iff $K, w \Vdash \beta$ or $K, w \Vdash \gamma$

(iii) $K, w \Vdash \neg\beta$ iff $K, w \nVdash \beta$

(iv) $K, w \Vdash \Box\beta$ iff for every $t \in W$, $K, t \Vdash \beta$

DEFINITION 176. If K is a Kripke structure, then K is a *model* of Γ if for every γ there is a world $w \in W$ such that $K, w \Vdash \gamma$.

THEOREM 177.

(1) Γ *has a model iff it is nontrivial.*

(2) *There are inconsistent sets of formulas that have models.*

Proof. Immediate application of Kripke semantics for $\mathcal{J}$. ■

These results reinforce that $\mathcal{J}$ is a paraconsistent logic.

We can define first-order discussive calculi with necessary identity as follows (these calculi can also be extended to higher-order systems). Let us call $\mathcal{J}^*$ the calculus whose language is that of the quantificational system S5Q$^=$ of Hughes and Creswell (see [Hughes and Cresswell, 1968]). We then define:

$$\Gamma \models_{\mathcal{J}^*} \alpha \text{ iff } \Diamond\Gamma \models_{S5Q^=} \Diamond\alpha.$$

In particular, $\models_{\mathcal{J}} \alpha$ iff $\models_{S5Q^=} \Diamond\alpha$. Given these definitions, all the results above can be applied to $\mathcal{J}^*$.

It is possible to present a sound and complete axiomatization for $\mathcal{J}^*$ as follows:

(J1*) If α is an axiom of S5Q$^=$, then $\Box\alpha$.

(J2*) Rules similar to (J2) to (J5) of $\mathcal{J}$ (see page 844).

(J3*) $\Box(\alpha \to \beta(x)) / \Box(\alpha \to \forall x\beta(x))$, where x is not free in α.

It is easy to see that in $\mathcal{J}^*$, $\to_d$, $\wedge_d$, $\vee_d$, $\leftrightarrow_d$, ∇, $\forall$ and $\exists$ have all the classical properties of $\to$, $\wedge$, $\vee$, $\leftrightarrow$, $\neg$, $\forall$ and $\exists$, respectively. Furthermore, it is not difficult to justify that: (i) S5Q$^=$ is contained in $\mathcal{J}^*$; (ii) both $\mathcal{J}^*$ and S5Q$^=$ are not decidable; (iii) both have algebraic semantics in which they are sound and complete; (iv) there are inconsistent, but nontrivial sets of formulas of $\mathcal{J}^*$, and finally, (v) a set of formulas Γ is nontrivial iff there is a Kripke structure for $\mathcal{J}^*$ that is a model of Γ. As a result, there are inconsistent sets of formulas that do have models.

There is another, and equivalent, way of building a first-order discussive calculus. This is suitable for certain applications, as will become clear below. We will call this calculus $\mathcal{J}^{**}$ (it is essentially the same as $\mathcal{J}^*$). The language of $\mathcal{J}^{**}$ is also the same as that of S5Q$^=$, except that if α is a formula, we write $\uplus\alpha$ to denote the formula obtained by preceding α by a sequence of universal quantifiers so that all variables of $\uplus\alpha$ are bound.

To define the calculus $\mathcal{J}^{**}$, we stipulate that

$$\vdash_{\mathcal{J}^{**}} \alpha \text{ iff } \vdash_{S5Q^=} \Diamond \uplus \alpha.$$

Furthermore, we say that α is a syntactic consequence of Γ in $\mathcal{J}^{**}$, that is, $\vdash_{\mathcal{J}^{**}} \alpha$, iff there are $\gamma_1, \ldots, \gamma_n \in \Gamma$ such that $\Diamond \uplus (\gamma_1 \wedge \ldots \wedge \gamma_n \to \alpha)$ is valid in S5Q$^=$. The postulates of $\mathcal{J}^{**}$ are (with the usual restrictions):

(J1**) If α is an instance of a classical tautology, then $\Box \uplus \alpha$ is an axiom.

(J2**) $\Box \uplus \alpha, \Box \uplus (\alpha \to \beta) / \Box \uplus \beta$

(J3**) $\Box \uplus (\Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta))$

(J4**) $\Box \uplus (\Box\alpha \to \alpha)$

(J5**) $\Box \uplus (\Diamond\alpha \to \Box\Diamond\alpha)$

(J6**) $\Box \uplus (\forall x\alpha(x) \rightarrow \alpha(t))$

(J7**) $\Box \uplus \alpha / \alpha$

(J8**) $\Box \uplus \alpha / \Box \uplus \Box\alpha$

(J9**) $\Diamond \uplus \alpha / \alpha$

(J10**) $\Box \uplus (\alpha \rightarrow \beta(x)) / \Box \uplus (\alpha \rightarrow \forall x\beta(x))$

(J11**) Vacuous quantification may be introduced in any formula.

(J12**) $\Box \uplus (x = x)$

(J13**) $\Box \uplus (x = y \rightarrow (\alpha(x) \rightarrow \alpha(y)))$

THEOREM 178. *In $\mathcal{J}^{**}$, the connectives $\rightarrow_d$, $\wedge_d$, $\vee_d$, ∇, $\leftrightarrow_d$, $\forall$ and $\exists$ have all the standard properties of classical (material) implication, conjunction, disjunction, negation, (material) equivalence, and of the universal and existential quantifiers, respectively.*

Thus, classical first-order logic is contained in $\mathcal{J}^{**}$. Furthermore, when we restrict the formulas of $\mathcal{J}^{**}$ to the *stable* ones only — that is, to those formulas α such that $\Box(\alpha \leftrightarrow_d \Diamond\alpha)$ is true — then $\mathcal{J}^{**}$ reduces, in a certain sense, to classical first-order logic.

4.2 Application to the foundational analysis of physical theories

A physical theory T can be characterized roughly as follows. We start with a language L by means of which we express the postulates of T. These postulates, or axioms, can be divided up into three kinds: the logical axioms (say, those of first-order predicate calculus), the mathematical axioms (say, those of a set theory like Zermelo-Fraenkel), and the specific, or physical, axioms of T. It is clear that this schema is quite general, and can be suitably adapted to several contexts (see [da Costa and French, 2003]). In particular, this schema is useful to an analysis of physical theories when we consider the possibility of inconsistencies. For instance, in certain situations, it may be more adequate to use a calculus such as $\mathcal{J}^{**}$ instead of standard first-order logic. This is the case we would like to consider here.

According to Dalla Chiara and Toraldo di Francia (see [Dalla Chiara and di Francia, 1981]), the language of a physical theory T is interpreted in a set-theoretic structure of the form

$$\mathcal{A} = \langle M, S, \langle Q_0, \ldots, Q_n \rangle, \rho \rangle,$$

where

(1) M is an instance of a mathematical species of structure in the sense of Suppes-Bourbaki (see [da Costa and Chuaqui, 1988]).

(2) S is a set of physical situations; that is, a set of physical states assumed by a physical system in a certain interval of time.

(3) Each Q_i, $0 \leq i \leq n$, denotes an operationally defined quantity whose domain of definition is some subset of S. In general, Q_0 represents time.

(4) ρ is a function that associates to each term employed to characterize S, and in particular to each Q_i, a set-theoretic entity in M.

When we measure a physical quantity Q_i of a physical system in a state $s \in S$ in a certain time t_j, we usually consider that its 'acceptable values' $q_i(t_j)$ lie in an interval $[q_i - \epsilon, q_i + \epsilon]$ of the real number line. (The length ϵ depends on the specific measurement technique and on the nature of the quantity in question.) In a certain sense, *all values* in the interval are 'appropriate values' for the measurement of the quantity Q_i of the physical system in a state $s \in S$. For instance, in measuring the table where we are working just now, we should accept (for all 'practical purposes') any value in the interval 1.20 ± 10^{-3} meters.

Let $\alpha(t, q_i(t_j))$ be a formula of L whose only free variables are t and t_j (which stand for time). Dalla Chiara and Toraldo di Francia consider the case of partial formulas, that is, formulas that are not defined for all values of their variables and parameters. But here we will only discuss 'total' formulas. We say that $\alpha(t, q_i(t_j))$ is *true* with respect to a situation s, written $\vDash_s \alpha$, if there are values t^0 of t in the considered time interval and q_i^0 of Q_i in the corresponding interval, such that $\alpha(t^0, q_i^0)$ is true in M in the standard (Tarskian) sense. We also say that $\alpha(t, q_i(t_j))$ is true in $\mathcal{A}$, and write $\mathcal{A} \models \alpha(t, q_i(t_j))$, if $\alpha(t, q_i(t_j))$ is true in M for every $s \in S$. Paraconsistency enters in Dalla Chiara and Toraldo di Francia's approach whenever we get $\overline{t^0}$ and $\overline{q_i^0}$ also in the intervals such that $\neg\alpha(\overline{t^0}, \overline{q_i^0})$ is also true in M. Let us explain this case with an example.

Consider Newton's second law $f = m.a$. The three physical variables appearing in this equation correspond to three deterministic physical quantities: *force* (f), *mass* (m) and *acceleration* (a), which are the physical quantities to be measured. Their acceptable range of values for a certain physical situation s lie, respectively, within three intervals: $[f_1, f_2] \subseteq \mathbb{R}$, $[m_1, m_2], and \subseteq \mathbb{R}$ $[a_1, a_2] \subseteq \mathbb{R}$, each expressing a certain precision ϵ for the measurements. It is the case that $\vDash_s f = m.a$ whenever there are three real numbers $p_1 \in [f_1, f_2]$, $q_1 \in [m_1, m_2]$ and $r_1 \in [a_1, a_2]$, such that $p_1 = q_1.r_1$.

However, due to the imprecision ϵ, there are also other three real numbers p_2, q_2 and r_2, each in the respective interval, so that $p_2 \neq q_2.r_2$, and these numbers are also *acceptable values* for the measurements of the corresponding physical quantities. So, strictly speaking, it is also the case that $\vDash_s \neg(F = m.a)$. In other words, the negation of Newton's law should also be true in the same physical situation s. Hence, we may have, for a sentence α and physical situation $s \in E$, both $\vDash_s \alpha$ and $\vDash_s \neg\alpha$. However, it is not the case that $\vDash_s \alpha \wedge \neg\alpha$, for this last case would entail the existence of three real numbers p', q' and r', belonging to the respective intervals, such that $\vDash_s p' = q'.r' \wedge p' \neq q'.r'$, which is impossible (see [Dalla Chiara and di Francia, 1981, p. 66]).

This definition of truth reflects a kind of *empirical truth*. And it has interesting consequences, as Dalla Chiara and Toraldo di Francia point out. For example, the logical connectives are not truth-functional, in the sense that the truth of a conjunction is not equivalent to the simultaneous truth of both conjuncts. Note also the 'paraconsistent aspect' of this definition of truth. After all, we can have both $\vDash_s \alpha$ and $\vDash_s \neg\alpha$. In specifying the underlying logic of Dalla Chiara and Toraldo di Francia's approach, we may use the postulates of a Jaśkowski's discussive logic, such as $\mathcal{J}^{**}$, as logical axioms for such physical theories. (Alternatively, we could use a paraclassic logic instead; see section 7.3.)

4.3 Application to partial truth

In a paper by Mikenberg, da Costa and Chuaqui (see [Mikenberg *et al.*, 1986]), the mathematical concept of pragmatic truth (today called 'partial truth') was introduced. An infinitary logical system was presented to formulate this concept, and some applications were made in logic and in algebra. An application of this concept in the foundations of the theory of probability was studied in [da Costa, 1986b], and some of its extensions to inductive logic and to the philosophy of science were discussed, respectively, in [da Costa and French, 1989a] and [da Costa and French, 1990]. This framework was also employed to examine some issues involved in the theory of acceptance (see [da Costa and French, 2003a]), as well as in the modeling of 'natural reasoning' (see [da Costa and French, 1993b]). For a general discussion, see [da Costa and French, 2003].

The wide range of applications of the notion of partial truth motivates an investigation of the logic of partial truth. In this section, we will show that there are important connections between this logic and Jaśkowski's discussive logic. We will study these connections in the context of certain 'pragmatic structures', which we will present below. As will become clear, pragmatic structures can be considered as worlds of a Kripke structure, and in this setting, the necessity operator corresponds to the notion of 'pragmatic validity', and the possibility operator to the notion of partial truth. Two systems are then put forward to formalize these notions. One of the main results of the present section is that the logic of partial truth is paraconsistent. The philosophical significance of this result, which justifies the application of partial truth to inconsistent settings, is then discussed.

A remark on our terminology is important here. We call the kind of truth defined in this section *partial truth*. Originally, it was called 'pragmatic truth', due to its connections with the pragmatic conception of truth, as developed by philosophers such as James, Dewey and particularly Peirce (see [Mikenberg *et al.*, 1986], [da Costa, 1986b], and [da Costa and French, 2003]). However, our work is not exegetical. The only point to emphasize is that our definition was heuristically inspired by some passages of pragmatic thinkers, such as Pierce, when he wrote: 'consider what effects, that might conceivably have practical bearings, we conceive the object of our conceptions to have. Then, our conception of these effects is the whole of our conception of the object' ([Peirce, 1965, p. 31]). In our view, the definition of partial, or pragmatic, truth investigated in this section captures, at least in part, the common concept of a theory *saving the appearances*, usually by means of partially fictitious constructions (see [Vaihinger, 1952] and [Bueno, 1997]). We now move on to the formulation of partial truth.

Let us suppose that we are interested in studying a certain domain of knowledge Δ in the field of empirical sciences, for instance, particle physics. We are, then, concerned with certain real objects (in particle physics, with some configurations in a Wilson chamber, some spectral lines, etc.). Let us denote the set of these objects by A_1. Among the objects of A_1, there are some relations that interest us, and that we model as partial relations R_i,

$i \in I$ (every relation having a fixed arity). The relations R_i are partial relations, that is, each R_i, supposed of arity r_i, is not necessarily defined for all r_i-tuples of elements of A_1. More formally, an n-place partial relation R can be viewed as a triple $\langle R_1, R_2, R_3 \rangle$, where R_1, R_2, and R_3 are mutually disjoint sets, with $R_1 \cup R_2 \cup R_3 = D^n$, and such that R_1 is the set of n-tuples that (we know) belong to R; R_2 the set of n-tuples that (we know) do not belong to R; and finally R_3 of those n-tuples for which it is not defined whether they belong or not to R. (Note that when R_3 is empty, R is a usual n-place relation that can be identified with R_1.)

The reason for using partial relations is that they are supposed to express what we do know, or what we accept as true, about the actual relations among the elements of A_1. Thus, the partial structure $\langle A_1, R_i \rangle_{i \in I}$ encompasses, so to say, what we know, or accept as true, about the actual structure of Δ. However, to systematize our knowledge of Δ, it is convenient to introduce in our structure $\langle A_1, R_i \rangle_{i \in I}$ some *ideal* objects. (In particle physics, quarks would be an example.) The set of these new objects will be denoted by A_2. It is understood that $A_1 \cap A_2 = \emptyset$, and we stipulate that $A = A_1 \cup A_2$. In this way, the modeling of Δ involves new partial relations R_j, $j \in J$, some of which extend the relations R_i, $i \in I$. Furthermore , there are some sentences (closed formulas) of the language L, in which we talk about the structure $\langle A, R_k \rangle_{k \in I \cup J}$ $(I \cap J = \emptyset)$ that we accept as true, or that are true (in the sense of the correspondence theory of truth). This occurs, for instance, with sentences expressing true decidable propositions (a proposition whose truth or falsehood can be *decided*), and with some general sentences that express laws or theories already accepted as true. Let us denote the set of such sentences, dubbed *primary*, by $\mathcal{P}$ (this set may be empty).

Given this informal discussion, we suggest that what we call a *simple pragmatic structure* (sps) be regarded as a set-theoretic structure of the form:

$$\mathfrak{A} = \langle A_1, A_2, R_i, R_j, \mathcal{P} \rangle_{i \in I, j \in J},$$

where the elements in question satisfy the conditions above. Alternatively, we can simply write:

$$\mathfrak{A} = \langle A, R_k, \mathcal{P} \rangle_{k \in K}$$

for a sps, where $A = A_1 \cup A_2$ and the R_k are partial relations defined on A, and $\mathcal{P}$ is a set of sentences of the language L of the same similarity type as that of $\mathfrak{A}$, and which is interpreted in $\mathfrak{A}$. Note that for some k, R_k may be empty.

Let L be a first-order language with identity, but without function symbols. The symbols of L are logical symbols (connectives, individual variables, quantifiers, and the identity symbol), auxiliary symbols (parentheses), a collection of individual constants, and a collection of predicate symbols. To interpret L in a sps $\mathfrak{A}$ is to associate to each individual constant of L an element of A (the universe of $\mathfrak{A}$), and to each n-ary predicate symbol of L a relation R_k, $k \in K$, of the same arity. It is supposed that every predicate of the family R_k, $k \in K$, is associated with a predicate symbol.

DEFINITION 179. Let L and $\mathfrak{A} = \langle A, R_k, \mathcal{P} \rangle_{k \in K}$ be, respectively, a language and a sps in which L is interpreted. Let $\mathfrak{B}$ be a *total* structure, that is, a usual structure whose n-ary relations are defined for all n-tuples of elements of its universe. And suppose that L is also interpreted in $\mathfrak{B}$. Then, $\mathfrak{B}$ is said to be $\mathfrak{A}$-*normal* if the following conditions are met:

(1) *The universe of* $\mathfrak{B}$ is A.

(2) *The (total) relations of* $\mathfrak{B}$ *extend the corresponding partial relations of* $\mathfrak{A}$.

(3) *If c is an individual constant of L, then in both* $\mathfrak{A}$ *and* $\mathfrak{B}$, *c is interpreted by the same element.*

(4) *If* $\alpha \in \mathcal{P}$, *then* $\mathfrak{B} \models \alpha$.

Given a pragmatic structure $\mathfrak{A}$, it may happen that there are no $\mathfrak{A}$-normal structures. It is possible, however, to provide a system of necessary and sufficient conditions for the existence of such structures (see [Mikenberg *et al.*, 1986]). One condition of this system is the following: For each partial relation R_k in $\mathfrak{A}$, we construct a set M_k of atomic sentences and negations of atomic sentences such that the former correspond to n-tuples that satisfy R_k, and the latter to n-tuples that do not satisfy R_k (such sentences correspond to n-tuples in the 'anti-extension' of R_k). Let M be the set $\bigcup_{k \in K} M_k$. Therefore, a sps $\mathfrak{A}$ admits an $\mathfrak{A}$-normal structure only if the set $M \cup \mathcal{P}$ is consistent.

In what follows, we will always suppose that our sps satisfies the relevant conditions; in other words, given any sps $\mathfrak{A}$, the set of $\mathfrak{A}$-normal structures is not empty.

DEFINITION 180. Let L and $\mathfrak{A}$ be, respectively, a language and a sps in which L is interpreted. We say that a sentence α of L is *pragmatically true*, or *partially true* in the sps $\mathfrak{A}$, according to $\mathfrak{B}$, if

(1) $\mathfrak{B}$ *is an* $\mathfrak{A}$*-normal structure, and*

(2) $\mathfrak{B} \models \alpha$, *that is,* α *is true in* $\mathfrak{B}$ *in accordance with the Tarskian definition of truth.*

In other words, we say that α is *pragmatically (or partially) true* in the sps $\mathfrak{A}$ if there exists an $\mathfrak{A}$-normal structure $\mathfrak{B}$ in which α is true in the standard Tarskian sense. If α is not pragmatically (partially) true in the sps $\mathfrak{A}$ according to $\mathfrak{B}$ (α is not pragmatically (partially) true in the sps $\mathfrak{A}$), we say that α is *pragmatically (partially) false* in the sps $\mathfrak{A}$ according to $\mathfrak{B}$ (α is pragmatically (partially) false in the sps $\mathfrak{A}$).

Given a sps $\mathfrak{A}$, it is natural to consider its $\mathfrak{A}$-normal structures as the worlds of a Kripke structure for S5 with quantification. That is, we have a universe and several structures, defined in this universe, in which the language L can be interpreted, and where every world is accessible to every world (see [Hughes and Cresswell, 1968]). It is also natural to extend the language L of the sps $\mathfrak{A}$ to a modal language, by adding the modal operator $\Box$ to the primitive symbols of L. The operator $\Box$, which in modal logic represents the notion of necessity, corresponds in the present situation to *pragmatic validity* (in a sps $\mathfrak{A}$). Analogously, the possibility symbol $\Diamond$, definable in terms of $\Box$ and negation, corresponds to *pragmatic truth* (in a sps $\mathfrak{A}$). Thus, we can extend the semantics of L so that the symbols $\Box$ and $\Diamond$ represent the concepts of pragmatic validity and pragmatic truth, respectively. Moreover, since the universes of all 'worlds' belonging to a sps are the same, it is reasonable that identity behaves, in the cases of pragmatic truth and pragmatic validity, as necessary identity.

Among the pragmatically valid formulas — that is, those formulas α such that $\Box\alpha$ is a theorem of S5 with quantification and necessary identity — there are the logically pragmatically true formulas — that is, those formulas α such that $\Box\Diamond\alpha$ or, equivalently, $\Diamond\alpha$ is a theorem of the same system. To simplify the language, from now on, the former class of formulas will be called *strictly pragmatically valid* and the latter will be called *pragmatically valid.* The first class of formulas coincides with the set of theorems of S5 with quantification and necessary identity; the second, with Jaśkowski's logic associated with the same system.

We first present the logical system that systematizes the notion of strict pragmatic validity. It will be denoted by *PV*. The system has a language L^* whose primitive symbols are those of a standard formalization of the first-order predicate calculus with identity and individual constants, plus the symbol $\Box$. (For simplicity, function symbols are excluded.) The defined symbols are introduced as usual, and the common conventions in the writing of formulas and in the formulation of postulates (axiom schemas and primitive rules of inference) are employed without explicit mention. The postulates of *PV* are the following:

(1) If α is an instance of a (propositional) tautology, then α is an axiom.

(2) $\alpha, \alpha \rightarrow \beta / \beta$

(3) $\Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$

(4) $\Box\alpha \rightarrow \alpha$

(5) $\Diamond\alpha \rightarrow \Box\Diamond\alpha$

(6) $\forall x\alpha(x) \rightarrow \alpha(t)$

(7) $\alpha / \Box\alpha$

(8) $\alpha \rightarrow \beta(x) / \alpha \rightarrow \forall x\beta(x)$

(9) $x = x$

(10) $x = y \rightarrow (\alpha(x) \rightarrow \alpha(y))$

In the postulates above, the symbols have a clear meaning. In particular, in the axiom schema (6), t is either a variable free for x in $\alpha(x)$ or an individual constant. This system is essentially S5 with quantification and necessary identity.

We define the concept of deduction as in [Henkin and Montague, 1956]. The basic idea is that one can only use the generalization rule $\alpha/\forall\alpha$ in a step k of a deduction when a subsequence of the deduction up to k is a proof of α. This restriction to the generalization rule is exactly similar to the one adopted with regard to the necessitation rule, $\alpha/\Box\alpha$. As a result, the usual derived rules, such as the deduction theorem, remain valid.

The semantics of *PV* can be easily developed. The basic (strict) semantic concepts of pragmatic truth, pragmatic falsehood, pragmatic validity, pragmatic invalidity, pragmatic semantic consequence, etc. offer no difficulties in being formulated, and maintain the spirit of Definition 180. (For more details on these points, see [da Costa *et al.*, 1998b] and [da Costa *et al.*, 1996].)

We have the following theorem, whose proof can be obtained by the methods of [Hughes and Cresswell, 1968] and [Gallin, 1975]:

THEOREM 181. *Let* Γ *be a set of formulas of* L^* *and* α *be a formula of the same language. Then,* $\Gamma \vdash \alpha$ *if, and only if,* $\Gamma \models \alpha$.

That is, α is a *syntactic consequence* of Γ in PV iff α is a *strict pragmatic semantic consequence* of Γ.

The logic of strict pragmatic validity, which we have just sketched, can be extended to higher-order (modal) languages, for example by adapting some ideas presented in [Gallin, 1975, Chap. 3]. Moreover, we can also develop a metatheoretic study of this logic. For instance, we can adopt different modal systems as basic (S4, for instance), distinguish frames from models etc. Instead of pursuing this path here, we will now consider a different system to study the logic of pragmatic validity. This system, as we will see, is constructed in terms of PV, which was presented here mainly as an auxiliary construction. We will then show that the logic of pragmatic validity is paraconsistent.

Let us call PT (for 'pragmatic truth') a system whose language is the same as that of PV. The idea is to construct a system in which $\vdash \alpha$ means that $\Diamond\alpha$ is strictly pragmatically valid. As before, if α is a formula of L^*, which is the language of PV and PT, we write $\uplus\alpha$ to denote the formula composed by preceding α by a sequence of universal quantifiers, so that all variables of $\uplus\alpha$ are bound.

Clearly, in order that α be pragmatically valid in the intended sense, we must have that $\vdash \alpha$ in PT if, and only if, $\vdash \Diamond \uplus \alpha$ in PV. So, PT is a kind of Jaśkowski's discussive logic associated with PV. (Recall that, given a modal system M, the Jaśkowski's logic associated with M is the set of all formulas α such that $\Diamond\alpha$ is a theorem of M.)

PT can be axiomatized as $\mathcal{J}^{**}$ of the previous section (see postulates J1** to J13** of page 848). The definitions of proof and of (formal) theorem are the usual ones. We now show that that postulates do provide an axiomatization for PT.

LEMMA 182. *If* α *is a theorem of our proposed axiomatization for* PT*, then* $\Diamond \uplus \alpha$ *is a theorem of* PV.

Proof. By induction on the length of the proof of α in the proposed axiomatization for PT. Let $\alpha_1, \ldots, \alpha_n$, where α_n is α, be a (formal) proof of α in the proposed axiomatization for PT. Then, α_i, $1 \leq i \leq n$, is an axiom or is obtained by the application of one of the rules. If a_i is an axiom, then it has the form $\Box \uplus \beta$, where β is an axiom of PV (observe that PV is S5 with quantification and necessary identity). Therefore, $\Box \uplus \beta$ is a theorem of PV, and so $\Diamond \uplus \Box \uplus \beta$, i.e. $\Diamond \uplus \alpha_i$, is also a theorem of PV. Suppose that α_i is a consequence of two previous formulas by Rule J2**. Then α_i is $\Box \uplus \beta$, obtained from the premises $\Box \uplus \gamma$ and $\Box \uplus (\gamma \to \beta)$. By the induction hypothesis, $\Diamond \uplus \Box \uplus \gamma$ and $\Diamond \uplus \Box \uplus (\gamma \to \beta)$ are provable in PV. Consequently, $\Box \uplus \gamma$ and $\Box \uplus (\gamma \to \beta)$ are also provable in PV, and so is $\Box \uplus \beta$. But if $\Box \uplus \beta$ is a theorem of PV, then $\Diamond \uplus \Box \uplus \beta$, i.e. $\Diamond \uplus \alpha_i$, is also a theorem. The other rules are treated similarly. ■

LEMMA 183. *If α is a theorem of PV, then $\Box \uplus \alpha$ is a theorem of the proposed axiomatization for PT.*

Proof. By induction on the length of the proof of α in PV. Let $\alpha_1, \dots, \alpha_n$, where α_n is α, be a proof of α in PV. If α_i, $1 \leq i \leq n$, is an axiom of PV, then $\Box \uplus \alpha$ a is a theorem of the proposed axiomatization for PT, as is easy to see. If α_i is obtained by an application of modus ponens (Rule J2**), from γ and $\gamma \to \alpha_i$, we have, by the induction hypothesis, that $\Box \uplus \gamma$ and $\Box \uplus (\gamma \to \alpha_i)$ are provable in the proposed axiomatization. Then, by Rule J2**, $\Box \uplus \alpha_i$ is also provable. Rule J8** is treated analogously. ■

THEOREM 184. *Postulates J1**–J13** characterize PT. That is, we have:*

$$\vdash \alpha \text{ in } PT \text{ iff } \vdash \Diamond \uplus \alpha \text{ in } PV.$$

Proof. Let us suppose that α is a theorem of the proposed axiomatization for PT. Then, by Lemma 182, $\Diamond \uplus \alpha$ is a theorem of PV. Conversely, assume that $\Diamond \uplus \alpha$ is a theorem of PV. So, by Lemma 183, $\Box \uplus \Diamond \uplus \alpha$ is a theorem of the proposed axiomatization for PT. Therefore, by Rule J7**, $\Diamond \uplus \alpha$ is a theorem of PT, and so, by Rule J9**, α is also a theorem of PT. ■

DEFINITION 185. In PT, we say that the formula α is a syntactic consequence of a set of formulas Γ (in symbols, $\Gamma \vdash \alpha$) if there exist $\gamma_1, \dots, \gamma_n$ in Γ such that

$$(\Diamond\gamma_1 \wedge \dots \wedge \Diamond\gamma_n) \to \Diamond\alpha$$

is a theorem of PT. Or, equivalently,

$$\Diamond \uplus ((\Diamond\gamma_1 \wedge \dots \wedge \Diamond\gamma_n) \to \Diamond\alpha)$$

is a theorem of PV. When $n = 0$, the first formula above reduces, by convention, to α (and $\emptyset \vdash \alpha$ means, thus, that $\vdash \alpha$).

DEFINITION 186. A *pragmatic theory* is a set T of sentences (closed formulas of PT) such that if $\gamma_1, \dots, \gamma_n$ are in T and $\{\gamma_1, \dots, \gamma_n\} \vdash \alpha$, then α is also in T.

It follows that if T is a pragmatic theory and α is a (closed) theorem of PT, then $\alpha \in T$. Let $\mathcal{S}$ be the set of all sentences of PT and T be a pragmatic theory. Using a terminology that has already been introduced, we say that T is *trivial* (*overcomplete*) if $T = \mathcal{S}$; otherwise, T is *nontrivial.* Furthermore, T is *inconsistent* if there is at least one sentence α such that $\alpha \in T$ and $\neg\alpha \in T$, where $\neg$ is the symbol of negation of PT; otherwise, T is *consistent.* We can now prove the following result:

THEOREM 187. *There exist pragmatic theories that are inconsistent but nontrivial.*

Proof. Let c and M be, respectively, an individual constant and a monadic predicate symbol of PT. The theory whose (nonlogical) axioms are $M(c)$ and $\neg M(c)$ is inconsistent. But it is nontrivial, because the corresponding theory of PV, whose (nonlogical) axioms are $\Diamond M(c)$ and $\Diamond\neg M(c)$, is consistent. In effect, it is easy to construct a Kripke model for PV in which both $\Diamond M(c)$ and $\Diamond\neg M(c)$ are true. However, in no Kripke model for PV the formula $\Diamond(M(c) \wedge \neg M(c))$ is true. ■

Given the definitions of the discussive connectives $\rightarrow_d$ and $\wedge_d$, introduced in Definition 170, we can also prove that, in *PT*, $\rightarrow_d$, $\wedge_d$, $\vee_d$, $\forall$ and $\exists$ satisfy all the schemas and rules of classical positive logic. If we consider a valid primitive schema (or rule) of classical positive logic, and replace in it implication by discussive implication and conjunction by discussive conjunction, we obtain a valid schema (or rule) of *PT*, as is easily seen.

THEOREM 188. *If T is a pragmatic theory, then $\alpha \in T$ iff there exist $\gamma_1, \ldots, \gamma_n$ in T such that*

$$(\gamma_1 \wedge_d \ldots \wedge_d \gamma_n) \rightarrow_d \Diamond\alpha$$

is a theorem of PT.

It is worth noting that in some applications of the theory just developed, it is sometimes convenient to employ an alternative definition of syntactic consequence. For instance, in certain applications in the foundations of physics, instead of Definition 185, it is more appropriate to adopt the following alternative:

DEFINITION 189. In *PT*, the sentence α is said to be a *proper syntactic consequence* of a set of sentences Γ if there are $\gamma_1, \ldots, \gamma_n$ in Γ such that $\Diamond(\gamma_1 \wedge \ldots \wedge \gamma_n)$ and $\Box((\gamma_1 \wedge \ldots \wedge \gamma_n) \rightarrow \alpha)$ are theorems of *PT* (or of *PT* and some extra axioms).

The philosophical significance of the above formal account can be seem by considering inconsistencies in our belief systems, as explored by da Costa, French and Bueno (see [da Costa *et al.*, 1998b]). If we focus on Theorem 187, it becomes clear that a pragmatic theory can contain contradictory theorems without becoming trivial. This means that *PT* belongs to the class of paraconsistent logics. In the case of pragmatic (partial) truth, this is not an unreasonable situation: contradictory propositions may, of course, both be pragmatically true (see [da Costa and French, 2003, Chap. 5]). Thus, partial truth can be used to provide the epistemic framework for characterizing inconsistent belief systems.

More precisely, we can formulate a position according to which 'belief that *p*' is not to be understood as 'belief that *p* is true' in the correspondence sense. When it comes to representational structures, such as scientific theories, 'belief that *p*' is to be understood as 'belief that *p* is pragmatically or partially true'. This allows for the accommodation of inconsistency by acknowledging that an inconsistency is not a permanent feature of reality to which theories correspond, but is rather a temporary aspect of these theories that may turn out to be epistemically fruitful. On this account, it is not the 'logic of science', in the sense of the underlying logic of deductive and inferential practices, that is paraconsistent, but rather the appropriate 'logic of truth'. (For further details, see [da Costa and French, 2003].)

The logic of pragmatic truth, as presented above, has also been developed to serve as a 'logic of scientific acceptance' (see [da Costa and French, 2003]). The nature of acceptance is a topic that hasn't received as much attention as it deserves. The accounts that consider the issue can be divided in two extremes: those that identify acceptance and belief, and those that distinguish the two. The former typically regard belief in terms of the correspondence conception of truth, whereas the latter fall prey to the accusation of some sort of anti-realism (or even conventionalism). An alternative is to retain the connection between belief and acceptance whilst rejecting truth-as-correspondence. On this view, to accept a theory is to be committed, not to believing it to be true *per se*, but to holding it *as if it were true*, for the purposes of further elaboration, development and investigation. Thus, acceptance involves belief that the theory is partially, or pragmatically, true only, and this, we believe, corresponds to the fallibility found in scientific practice.

Linking acceptance and pragmatic truth in this way restores a formal similarity between 'truth', taken generally, and acceptance with regard to deductive closure. So, it has been argued, for example, that acceptance differs from truth in that whereas the latter is deductively closed, in the sense that what one deduces from a set of truths is also true, the former generally is not (see [Ullian, 1990]). This is correct if closure is understood only in classical terms. However, what the above formal analysis shows is that acceptance, understood within the framework of pragmatic truth, may be regarded as closed under the Jaśkowski's discussive system. To put it more precisely: although there is no closure with regard to acceptance under classical conjunction and material implication, one can define discussive forms of implication and conjunction as above, with respect to which acceptance can be taken as closed.

This is a result of both general and particular significance. Our contention is that inconsistency can be accommodated in a framework in which accepted propositions are closed under implication. The framework is the one in terms of pragmatic truth, and the form of implication is, of course, discussive. Shifting perspective from the specific to the general, it is the failure to consider such nonclassical systems that undercuts the claim that 'logic' is not specially relevant to reasoning. Within the framework of pragmatic truth, we can accommodate inconsistency while still retaining a sense of deductive closure. In this manner the relevance of logic to reasoning – especially scientific reasoning — is restored (see [da Costa and French, 2003]).

5 ANNOTATED LOGICS

Reasoning about inconsistency is also important in computer science, data base theory, and artificial intelligence. For instance, to construct a knowledge base about a certain domain D of knowledge, we generally consult a certain number of experts, say $E_1, \ldots, E_n$, of that field of knowledge. Each expert contributes with facts and rules that form the bases $S_1, \ldots, S_n$ of sets of sentences. In this way, the whole knowledge base can be taken as the set $S_1 \cup \ldots \cup S_n$. However, experts may disagree, and often they do. As a result, this last set may be inconsistent, and so, according to classical first-order model theory, will have no models, and consequently, will be deemed 'meaningless'.

In 1987, H. Blair and V. S. Subrahmanian devised a kind of paraconsistent logic, called 'annotated logic', that was suitable for representing databases and knowledge bases that contain inconsistencies (see [Subrahmanian, 1987]). Later, Blair and Subrahamanian developed this framework further, endowing it with a fixed-point theory, a model theory and a proof theory (see [Blair and Subrahmanian, 1989c]). Moreover, they extended earlier results to allow for logic programming over a complete lattice of truth-values, extending accordingly the fixed-point theory and the proof theory (see [Blair and Subrahmanian, 1988b]). In turn, Kifer and Subrahmanian generalized annotated logic in such a way that a framework for logic programming based on the concept of 'bilattice' (which had been developed earlier by Fitting) was provided by the extended annotated logic framework (see [Kifer and Subrahmanian, 1989]). Kifer and Li then showed how annotated logic can be used as a foundation for reasoning in the presence of inconsistency (see [Kifer and Li, 1988]). And Kifer and Lozinskii demonstrated an embedding of classical logic in annotated logic; showed the connections between these logics and nonmonotonic logics, and devised a mechanical proof procedure for annotated logic (see [Kifer and Lozinskii, 1989]). Finally, Kifer and Wu showed how annotated logics serve as a foundation for object-oriented databases (see [Kifer and Wu, 1989]), whereas Kifer and Krishnaprasad showed how annotated logics can be used as a foundation for inheritance networks (see [Kifer and Kriphnaprasad, 1989]).

The foundational study of annotated logic was suggested by da Costa, Subrahmanian and Vago (see [da Costa *et al.*, 1991]), who developed a family of propositional calculi, called $\mathsf{P}\mathcal{T}$, as well as their first-order counterparts, $\mathsf{Q}\mathcal{T}$. This work has been extended to annotated set theory, and further results about $\mathsf{Q}\mathcal{T}$have been presented as well (see [da Costa *et al.*, 1991]).

5.1 *The annotated logic $\mathsf{Q}\mathcal{T}$*

$\mathsf{Q}\mathcal{T}$ is a first-order logic defined as follows. $\mathcal{T}$ is an arbitrary, but fixed, complete lattice. The least element of $\mathcal{T}$ is denoted by $\bot$, while the greatest element is denoted by $\top$. Furthermore, $\neg$ is taken to be a unary operator from $\mathcal{T}$ to $\mathcal{T}$. The language L of $\mathsf{Q}\mathcal{T}$ is a first-order language without identity whose primitive symbols are the following:

1. Connectives: $\rightarrow$ (implication), $\vee$ (disjunction), $\wedge$ (conjunction) and $\neg$ (negation).

2. Individual variables: a denumerably infinite set of variable symbols.

3. Individual constants: an arbitrary family of constant symbols.

4. Quantifiers: $\forall$ (universal quantifier) and $\exists$ (existential quantifier).

5. Function symbols: for each natural number $n > 0$, a collection of function symbols of rank n.

6. Predicate symbols: for each $n \geq 0$, a family of predicate symbols of rank n.

7. Auxiliary symbols: parentheses and comma.

Terms of L are introduced as usual. An (ordinary) *atom* is an expression of the form $P(t_1, \ldots, t_n)$. If P is a predicate symbol of rank n and $\lambda \in \mathcal{T}$, an *annotated predicate* is a pair $\langle P, \lambda \rangle$. We will denote an annotated predicate by P_λ, and sometimes will simply call it a 'predicate'.

Given an annotated predicate P_λ of rank n and n terms $t_1, \ldots, t_n$, an *annotated atom* is an expression of the form $P_\lambda(t_1, \ldots, t_n)$. The notion of *formula* is introduced in the standard way. Note that the symbol $\neg$ is used here in two distinct ways: first, as a mapping from $\mathcal{T}$ to $\mathcal{T}$, and second, as a unary connective of L. The appropriate meaning of $\neg$ will be given by the context.

DEFINITION 190. An *interpretation* I for the language L is a 4-tuple

$$I = \langle D, \eta_I, \zeta_I, \chi_I \rangle,$$

where:

(i) D is a non-empty set, called the *domain* of I.

(ii) η_I maps individual constants of L into D.

(iii) ζ_I assigns to each function symbol f of rank n a mapping from D^n to elements of D.

(iv) χ_I assigns to each predicate symbol P of rank n a function $\chi_I(P)$ from D^n to $\mathcal{T}$.

DEFINITION 191.

(i) Let I be an interpretation for L. Then a *variable assignment* v for L, with respect to I, is a map from the set of individual variable symbols of L to D.

(ii) The *denotation* $d_{I,v}(t)$ of a term t of L, with respect to an interpretation I and variable assignment v, is defined inductively as follows:

(a) If t is a constant symbol, then $d_{I,v}(t) = \eta_I(t)$.

(b) If t is an individual variable, then $d_{I,v}(t) = v_I(t)$.

(c) If t is $f(t_1, \ldots, t_n)$, then $d_{I,v}(t) = \chi_I(d_{I,v}(t_1), \ldots, d_{I,v}(t_n))$.

(iii) An annotated atom $P_\lambda(t_1, \ldots, t_n)$ will be denoted by $P(t_1, \ldots, t_n) : \lambda$. An expression of the form $\underbrace{\neg \ldots \neg}_{k\, times}(A : \mu)$, where A is an ordinary atom, is called a *hyper-literal* of order $k, k \geq 0$, and abbreviated by $\neg^k(A : \mu)$.

DEFINITION 192. *Let I be an interpretation for L, and let v be a variable assignment for L. Also, let A be an ordinary atom, and let α, β, and γ be arbitrary formulas. In this case:*

(i) If A is an ordinary atom $P(t_1, \ldots, t_n)$, then

$$I, v \models (A : \mu) \text{ iff } \chi_I(P)(d_{I,v}(t_1), \ldots, d_{I,v}(t_n)) \geq \mu.$$

(ii) $I, v \models \underbrace{\neg \ldots \neg}_{k\, times}(A : \mu)$ *iff* $I, v \models \underbrace{\neg \ldots \neg}_{k-1\, times}(A : \neg\mu)$.

It is important to note here the two senses in which $\neg$ is used. Inside the atom, that is, in $(A : \neg\mu)$, what appears is the map from $\mathcal{T}$ to $\mathcal{T}$. However, outside the atom, that is, in the first occurrence of $\neg$ in the expression $\neg(A : \neg\mu)$, it is the negation symbol that should be considered. Condition (ii) states that, to reduce negations, we replace values in the lattice according to the map $\neg$.

(iii) $I, v \models \alpha \wedge \beta$ *iff* $I, v \models \alpha$ *and* $I, v \models \beta$.

(iv) $I, v \models \alpha \vee \beta$ *iff* $I, v \models \alpha$ *or* $I, v \models \beta$.

(v) $I, v \models \alpha \to \beta$ *iff* $I, v \models \alpha$ *or* $I, v \not\models \beta$.

(vi) If F *is not a hyper-literal, then* $I, v \models \neg F$ *iff* $I, v \not\models F$.

(vii) $I, v \models \forall x\alpha$ *iff for every variable assignment* v' *that agrees with* v *in all variables distinct from* x *(that is,* $v'(y) = v(y)$*, for all* $y \neq x$*),* $I, v' \models \alpha$.

(viii) $I, v \models \exists x\alpha$ *iff for some variable assignment* v' *that agrees with* v *in all variables distinct from* x *(that is,* $v'(y) = v(y)$*, for all* $y \neq x$*),* $I, v' \models \alpha$.

(ix) $I \models \alpha$ *iff for all variable assignments* v *associated with* I*,* $I, v \models \alpha$.

DEFINITION 193.

1. Let $\Gamma \cup \{\alpha\}$ *be a set of formulas. We write* $\models \alpha$*, and say that* α *is* valid *(in* $\mathcal{QT}$*) if, for every interpretation* I*,* $I \models \alpha$*. We say that* α *is a* semantic consequence *of* Γ *iff for every interpretation* I *such that, for all* $\beta \in \Gamma$*,* $I \models \beta$*, it is the case that* $I \models \alpha$*.*

2. If α, β *are formulas of* L*, then* $\alpha \leftrightarrow \beta =_{\text{def}} (\alpha \to \beta) \wedge (\beta \to \alpha)$.

3. If α *is a formula, then* $\sim \alpha =_{\text{def}} (\alpha \to (\alpha \to \alpha)) \wedge \neg(\alpha \to \alpha)$.
($\sim \alpha$ *is called the* strong negation *of* α *in* $\mathcal{QT}$*.)*

4. A formula is called complex *if it is not a hyper-literal.*

We will describe now an axiomatic system, called $\mathcal{A}$, whose underlying language is L. In the postulates below, α, β and γ denote arbitrary formulas; φ and ψ denote complex formulas, and θ is an annotated atom.

The postulates of our system are:

$(\to_1)$ $\alpha \to (\beta \to \alpha)$

$(\to_2)$ $(\alpha \to \beta) \to ((\alpha \to (\beta \to \gamma)) \to (\alpha \to \gamma))$

$(\to_3)$ $\alpha,\ \alpha \to \beta / \beta$

$(\to_4)$ $((\alpha \to \beta) \to \alpha) \to \alpha$

$(\wedge_1)$ $\alpha \wedge \beta \to \alpha$

$(\wedge_2)$ $\alpha \wedge \beta \to \beta$

$(\wedge_3)$ $\alpha \to (\beta \to \alpha \wedge \beta)$

$(\vee_1)$ $\alpha \to (\alpha \vee \beta)$

$(\vee_2)$ $\beta \to (\alpha \vee \beta)$

($\vee_3$) $(\alpha \to \gamma) \to ((\beta \to \gamma) \to (\alpha \vee \beta \to \gamma))$

($\neg_1$) $(\varphi \to \psi) \to ((\varphi \to \neg\psi) \to \neg\varphi)$

($\neg_2$) $\varphi \to (\neg\varphi \to \alpha)$

($\neg_3$) $\varphi \vee \neg\varphi$

($\exists_1$) $\alpha(t) \to \exists x\alpha(x)$

($\exists_2$) $\frac{\alpha(x)\to\beta}{\exists x\alpha(x)\to\beta}$

($\forall_1$) $\forall x\alpha(x) \to \alpha(t)$

($\forall_2$) $\frac{\alpha\to\beta(x)}{\alpha\to\forall x\beta(x)}$

(τ_1) $(\theta :\perp) \wedge \neg^k(\theta : \mu) \leftrightarrow \neg^{k-1}(\theta : \neg\mu)$

(τ_2) $(\theta : \neg\mu) \to (\theta : \neg\lambda)$, where $\lambda \leq \mu$.

(τ_3) If $\alpha \to (\theta : \mu_j)$ for every $j \in J$, then $\alpha \to (\theta : \mu)$, where $\mu = \sup\{\mu_j : j \in J\}$.

The postulates ($\exists_1$)-($\forall_2$) satisfy the usual restrictions. If $\mathcal{T}$ is a complete lattice, the supremum in rule (τ_3) is well defined. When $\mathcal{T}$ is finite, rule (τ_3) can be replaced by the schema $(\theta : \mu_1) \wedge \ldots \wedge (\theta : \mu_n) \to (\theta : \mu)$, where $\mu = \sup\{\mu_j : 1 \leq j \leq n\}$.

We easily define the syntactic concepts associated to the axioms above. In particular, the concept of syntactic consequence $\vdash$ is defined as usual. However, the notion of deduction is not finitary if $\mathcal{T}$ is infinite. The propositional counterpart of Q$\mathcal{T}$ can also de developed, as shown in [da Costa *et al.*, 1991].

THEOREM 194 (Soundness). *Let* $\Gamma \cup \{A\}$ *be a set of formulas of* Q$\mathcal{T}$. *Then* $\Gamma \vdash \alpha$ *implies that* $\Gamma \models \alpha$. *That is, the axiomatic system is sound with respect to the semantics of* Q$\mathcal{T}$.

Proof. By induction on the length of deduction (if $\mathcal{T}$ is infinite, we use transfinite recursion). ■

DEFINITION 195. Suppose that Γ is a set of formulas such that the set of annotated constants occurring in Γ is finite (Γ itself may be infinite). In this case, Γ is said to have the *finite annotation property*.

We note that if $\mathcal{T}'$ is a substructure of $\mathcal{T}$, then $\mathcal{T}'$ is closed under the operations of $\mathcal{T}$.

THEOREM 196 (Finitary Completeness). *Let* $\Gamma \cup \{A\}$ *be a set of formulas of* Q$\mathcal{T}$. *Then if* $\mathcal{T}$ *is finite or if* $\Gamma \cup \{\alpha\}$ *has the finite annotated property, then* $\Gamma \models \alpha$ *entails* $\Gamma \vdash \alpha$.

Proof. By extending the proof of the propositional fragment of Q$\mathcal{T}$ presented in [da Costa *et al.*, 1991]. ■

When $\mathcal{T}$ is infinite, it seems that the completeness can be obtained by adding to the axioms an extra infinitary rule (see [da Costa *et al.*, 1991] for further indications). The system Q$\mathcal{T}$ is a nonclassical logic which is both paraconsistent and paracomplete (see [da Costa, 1997]).

Another axiomatization of $Q\mathcal{T}$.

A different axiomatization of $\mathsf{Q}\mathcal{T}$ can be obtained by adding to the language of classical first-order predicate calculus a symbol for the paraconsistent weak negation satisfying suitable axioms. As an example, let C be an axiomatic systematization of first-order predicate calculus without identity whose symbol of negation is $\sim$. The remaining primitive symbols of C are as those of $\mathsf{Q}\mathcal{T}$. We still suppose that the atomic formulas of C are annotated atoms, as above. Furthermore, our language has a primitive symbol $\neg$ for the weak negation.

Let us denote by $\mathcal{A}'$ the axiomatic system obtained from C by adding to it the axioms $(\neg_1)$, $(\neg_2)$, $(\neg_3)$, (τ_1), (τ_2) and (τ_3) plus the following rule:

[Rule] Let α and β be formulas such that β is obtained from α by replacing sub-formulas of the form $\sim \alpha$ by $\alpha \to ((\alpha \to \alpha) \wedge \neg(\alpha \to \alpha))$, or by replacing sub-formulas of the latter form by the first. We can then infer that $\alpha \leftrightarrow \beta$.

THEOREM 197. *The axiom system $\mathcal{A}$ and $\mathcal{A}'$ are equivalent. So, both characterize $Q\mathcal{T}$.*

Proof. Any postulate of $\mathcal{A}$ is a postulate of $\mathcal{A}'$, and item 3 of Definition 193 corresponds to a rule in $\mathcal{A}'$. Conversely, any postulate of $\mathcal{A}'$ is a postulate or a definition of $\mathcal{A}$, or is provable in $\mathcal{A}$, as is easy to show. ■

Let X be a non-empty set. A *normal structure based on X* is a function $f : X \times X \mapsto \mathcal{T}$. Let us denote by $\mathsf{Q}\mathcal{T}^2$ the logic obtained from $\mathsf{Q}\mathcal{T}$ by suppressing all function and predicate symbols from the language, with the exception of one binary predicate symbol, which we represent by $\in$. Then $\mathsf{Q}\mathcal{T}^2$ is a dyadic predicate logic whose atoms are annotated by $\mathcal{T}$. These atoms have the form $\in_\lambda (a, b)$, where a and b are terms, and $\lambda \in \mathcal{T}$. This atom will be denoted by $a \in_\lambda b$.

THEOREM 198. *$Q\mathcal{T}^2$ is sound with respect to the semantics of normal structures. If $\mathcal{T}$ is finite, or if we consider only sets of formulas having the finite annotation property, then $Q\mathcal{T}^2$ is also complete.*

Proof. Consequence of Theorems 194 and 196. ■

5.2 Annotated set theory

Theorem 198 above shows that normal structures are important to annotated logics, particularly when $\mathsf{Q}\mathcal{T}$ is developed along the lines of the previous section. In this section, we will extend annotated logic to set theory, and we will be dealing with normal structures throughout.

Let ZF be a standard formulation of the Zermelo-Fraenkel set theory. The language of annotated set theory, called AZF, is obtained from the language of ZF by adding two individual constants, $\mathcal{T}$ and $\mathcal{U}$. The following axioms are also added to those of ZF:

(AZF.1) $\mathcal{T}$ is a complete lattice, and we denote by $\leq$ an arbitrary, but fixed, ordering. We use $\perp$ and $\top$ to stand for its least and greatest elements, respectively.

(AZF.2) $\mathcal{T} \subseteq \mathcal{U}$ and $\forall x(x \in \mathcal{U} \to x \subseteq \mathcal{U})$. That is, $\mathcal{U}$ is transitive.

In most applications, it is usually enough to postulate that $\mathcal{T}$ is a set with a reflexive binary relation with unique least and greatest elements.

DEFINITION 199.

(1) We say that $\mathcal{E}$ is a normal structure, *or a* normal function *based on $\mathcal{U}$, if $\mathcal{E}$ is a mapping from $\mathcal{U} \times \mathcal{U}$ into $\mathcal{T}$. We write $x \in_\lambda y$ instead of $\mathcal{E}(x, y) = \lambda$.*

(2) If $x \in \mathcal{U}$, then:

(2.1) $x^{(\lambda,\mathcal{E})} =_{\mathsf{def}} \{y : y \in \mathcal{U} \wedge y \in_\lambda x\}$

(2.2) $x^{[\lambda,\mathcal{E}]} =_{\mathsf{def}} \{y : y \in \mathcal{U} \wedge \exists\mu(\mu \in \mathcal{T} \wedge \mu \leq \lambda \wedge y \in_\mu x)\}$

(2.3) $\mathcal{F}_x =_{\mathsf{def}} \{f : f : x \mapsto \mathcal{T} \wedge \exists\mathcal{E}(\mathcal{E}$ is a normal function $\wedge \forall\lambda\forall y(\lambda \in \mathcal{T} \wedge y \in \mathcal{U}) \to (f(y) = \lambda \leftrightarrow y \in_\lambda x)\}$.

(3) If $x, y \in \mathcal{U}$ and $\lambda \in \mathcal{T}$, then $x =_{\lambda,\mathcal{E}} y =_{\mathsf{def}} \forall z \in \mathcal{U}(z \in_\lambda x \leftrightarrow z \in_\lambda y)$.

(4) A set $x \neq \emptyset$ is strongly transitive *if it is transitive and $\forall y(y \in x \to \mathcal{P}(y) \in x)$, where $\mathcal{P}(y)$ is the power-set of y.*

(5) x is called a universe *iff it is strongly transitive and for every function $f : x \mapsto x$, if $y \in x$, then $\bigcup ran(f) \in x$, where $ran(f) =_{\mathsf{def}} \{z : \exists t(t \in y \wedge f(t) = z)\}$.*

THEOREM 200. *If x is a universe, then x is a standard model of all axioms of ZF, with the possible exception of the axiom of infinity. If $\omega \in x$, then x is a complete universe (ω is the set of natural numbers).*

Set theoretic constructions are used to deal with normal structures based on $\mathcal{U}$. It seems that the more such constructions are available, the better. In most cases, it is useful to take $\mathcal{U}$ as a universe, that is, as a model of ZF if $\omega \in \mathcal{U}$. In particular, we have the following result:

THEOREM 201. *If $\mathcal{U}$ is an universe, $\lambda \in \mathcal{T}$, and $x \in \mathcal{U}$, then $x^{[\lambda,\mathcal{E}]} \in \mathcal{U}$, $x^{(\lambda,\mathcal{E})} \in \mathcal{U}$, and $\mathcal{F}_x \in \mathcal{U}$. Furthermore, $\{y : F(y) \wedge y \in_\lambda x\} \in \mathcal{U}$, where $F(y)$ is any formula of ZF, and $\forall x\forall y(x, y \in \mathcal{U} \to (\forall\mathcal{E}\forall\lambda(\mathcal{E}$ is a normal function $\wedge \lambda \in \mathcal{T}) \to x =_{\mathcal{E},\lambda} y) \leftrightarrow x = y)$.*

Proof. Immediate, since $\mathcal{U}$ is a model of ZF (except for the axiom of infinity). ■

We can introduce a weak negation ¬ in AZF without difficulty. For instance, it applies to hyper-literals only, and is such that:

$$\neg^k(x \in_\lambda y) \leftrightarrow \neg^{k-1}(x \in_{\neg\lambda} y).$$

Fuzzy sets

The concept of a fuzzy set can be formulated in AZF. If $\mathcal{U}$ is a set, then a *fuzzy set* of $\mathcal{U}$ is a function $u : \mathcal{U} \mapsto [0,1]$. Let us denote by $\mathcal{F}_{\mathcal{U}}$ the set of all fuzzy sets of $\mathcal{U}$. We say that two fuzzy sets $u, v \in \mathcal{F}_{\mathcal{U}}$ are *equal* iff for every $x \in \mathcal{U}$, $u(x) = v(x)$. Let us use $\mathbf{1}_u$ and $\mathbf{0}_u$ to denote the fuzzy sets of $\mathcal{U}$ so that for every $x \in \mathcal{U}$, $\mathbf{1}_u(x) = 1$ and $\mathbf{0}_u(x) = 0$. Furthermore, if $u, v \in \mathcal{F}_{\mathcal{U}}$ and $x \in \mathcal{U}$, we stipulate that $(u \sqcup v)(x) =_{\mathsf{def}} \sup\{u(x), v(x)\}$; $(u \sqcap v)(x) =_{\mathsf{def}} \inf\{u(x), v(x)\}$, and $\bar{u} =_{\mathsf{def}} 1 - u(x)$.

It is then easy to prove that $\langle \mathcal{F}_{\mathcal{U}}, \sqcup, \sqcap \rangle$ is a complete lattice with the infinite distributive property, and that $\langle \mathcal{F}_{\mathcal{U}}, \sqcup, \sqcap, ^{-} \rangle$ is an algebra that, in general, is not Boolean. A fuzzy set u of $\mathcal{U}$ can be identified with a normal structure $\tilde{u}$ based on the set $\mathcal{U} \cup [0,1]$, such that $\mathcal{T} = \{\top, \bot\}$ and

$$\tilde{u} = \begin{cases} \top & \text{if } x \in \mathcal{U} \wedge y \in [0,1] \wedge y = u(x), \\ \bot & \text{otherwise.} \end{cases}$$

Hence, AZF, considered as the theory of normal structures, encompasses the theory of fuzzy sets. It is clear that if $\mathcal{U}$ is a universe, the definition of fuzzy set in terms of normal structures can be simplified. In the same way, the theory of flou sets and of L-sets (see [Negoita and Ralescu, 1975]) can also be obtained by extending the concept of normal structures. Further developments relating annotated logics and fuzzy logics can be found in [da Costa *et al.*, 1991] and [Abe, 1992].

5.3 Applications

Annotated logics have a wide range of applications. In this section, we will illustrate how annotated logics provide a formalism for reasoning about inconsistent knowledge bases.[15] (In section 7.2, we will sketch another use of annotated logics.) Although the technical details cannot be given here in complete detail, the references provide additional information.

Expert systems and knowledge bases about a domain D are usually constructed by programmers who, in general, know little about D. To build an expert system, say, in medicine, we need to consult several experts in the particular field we are interested in (say, cardiology). We ask the experts to provide us with adequate knowledge, based on their previous experience, so that we can construct a knowledge base. It is common to assume that the information provided by the experts can be expressed in suitable form in a certain logic language.

But experts usually disagree. For instance, given the same observable symptoms, doctor d_1 may believe that the patient has a virus infection, doctor d_2 may conclude that the patient has an allergic reaction, while doctor d_3 may say that the patient either has a viral infection or an allergy, but not both. It is clear that if we had used the opinions of these three doctors in our knowledge base, we would be led into an inconsistency. The important point is that often experts disagree, and have conflicting opinions, for very good reasons. So, inconsistencies, such as the one found in this sample, should be regarded as

---

[15] This section is partially based on [da Costa and Subrahmanian, 1989].

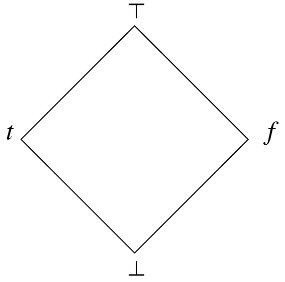


Figure 1. The Lattice FOUR

natural. In other words, to construct knowledge bases, we need to be sensitive to the fact that inconsistencies may be present, and take them seriously.

We present here a general framework for logic programming over a set of truth-values that has the structure of a complete lattice. We develop two formulations of the framework: one that provides a four-valued logic to reason in the presence of inconsistency, and another that shows how to reason with both inconsistency and uncertainty.

Let us assume that we have a fixed set $\mathcal{T}$ of truth-values that is a complete lattice under an ordering $\leq$ defined on $\mathcal{T}$. Let us denote the least upper bound and the greatest upper bound of the subsets $S \subset \mathcal{T}$, respectively, by $\sqcup S$ and $\sqcap S$. For instance, consider the lattice FOUR shown in Figure 1. Here, t and f represent the classical truth-values 'true' and 'false', respectively, while $\bot$ denotes 'unknown', and $\top$ stands for 'inconsistent' or 'over-defined'.

Intuitively, a lattice like FOUR can be useful to accommodate certain classically inconsistent theories. Consider, for example, the theory T axiomatized by $\{p, \neg p, q\}$. T has a model, namely, the interpretation that assigns $\top$ to p and t to q. Given the definition of satisfaction discussed below (see Definition 204), it can be shown that $\neg q$ is not a logical consequence of T. Of course, this is not the case in classical logic.

Another important complete lattice is SQUARE, that is, the set $SQ = [0,1] \times [0,1]$ of truth-values, where $[0,1] \subset \mathbb{R}$, endowed with the ordering below. Here, the assignment of a truth value $[\mu_1, \mu_2]$ to p means that the degree of belief in p is μ_1, while the degree of disbelief in p is μ_2. The ordering is the following:

$$[\mu_1, \mu_2] \leq [\rho_1, \rho_2] \text{ iff } \mu_1 \leq_{\mathbb{R}} \rho_1 \text{ and } \mu_2 \leq_{\mathbb{R}} \rho_2,$$

where $\leq_{\mathbb{R}}$ is the ordinary 'less than or equal to' defined on the real numbers. In SQ, $[0,0]$ intuitively denotes absolute lack of belief, $[0,1]$ denotes complete disbelief, $[1,0]$ denotes complete belief and $[1,1]$ denotes absolutely inconsistent beliefs.

Let $(A : \mu)$ be an annotated atom over $\mathcal{T}$ (see Definition 191). If α_1 and α_2 are first-order expressions (terms or atoms), then a *substitution* θ of variable symbols for terms is called a *unifier* of α_1 and α_2 iff the application of θ to α_1, denoted $\alpha_1\theta$, yields the same expression as $\alpha_2\theta$. A *most general unifier* (mgu for short) of any two syntactic expressions α_1 and α_2 is a unifier θ such that for any unifier ϑ of the expressions α_1 and α_2, there is a substitution γ such that $\theta\gamma = \vartheta$. If α_1 and α_2 are unifiable terms of atoms, then they posses a mgu (see [da Costa and Subrahmanian, 1989]).

DEFINITION 202. If $L_0, L_1, \ldots, L_n$ are annotated atoms over $\mathcal{T}$, then $L_0 \Leftarrow L_1 \wedge \ldots \wedge L_n$ is an *annotated clause* over $\mathcal{T}$. L_0 is called the head of the annotated clause, while $L_1 \wedge \ldots \wedge L_n$ is its body. We will often refer to annotated clauses just as clauses.

DEFINITION 203. An *annotated logic program* (ALP) over $\mathcal{T}$ is a finite set of annotated clauses over $\mathcal{T}$.

Semantics

Let us define a semantics for ALPs. To do that, we will consider only those interpretations whose domain of discourse are the set of ground terms of the language (Herbrand interpretations). An interpretation I of an ALP P over $\mathcal{T}$ is a mapping $I : Bp \mapsto \mathcal{T}$, where Bp is the Herbrand base of P, that is, the set of variable-free atoms expressible in the language of P. The ordering $\leq$ is extended to the interpretation in a natural way, namely:

$$I_1 \leq I_2 \text{ iff } (\forall A \in Bp)(I_1(A) \leq I_2(A)).$$

The orderings $\geq$, $<$ and $>$ are defined in the usual way. We also assume the existence of a function $\neg : \mathcal{T} \mapsto \mathcal{T}$.

DEFINITION 204 (Satisfaction). An interpretation I is said to satisfy

1. the formula α iff it satisfies every closed instance of α;
2. the variable-free annotated atom $(A : \mu)$ iff $I(A) \geq \mu$;
3. the variable-free annotated hyper-literal $\neg(A : \mu)$ iff $I(A) \geq \neg(u)$;
4. the variable-free formula $\alpha_1 \wedge \alpha_2$ iff I satisfies α_1 and α_2;
5. the variable-free formula $\alpha_1 \vee \alpha_2$ iff I satisfies α_1 or α_2;
6. the variable-free formula $\alpha_1 \Leftarrow \alpha_2$ iff either I satisfies α_1 or does not satisfy α_2;
7. the variable-free formula $\alpha_1 \Leftrightarrow \alpha_2$ (that is, $(\alpha_1 \Leftarrow \alpha_2) \wedge (\alpha_2 \Leftarrow \alpha_1)$) iff I satisfies $\alpha_1 \Leftarrow \alpha_2$ and $\alpha_2 \Leftarrow \alpha_1$;
8. the closed formula $\exists x\alpha$ iff there is some variable-free term t such that I satisfies $\alpha[x/t]$ (the result of replacing all free occurrences of x in α by t);
9. the closed formula $\forall x\alpha$ iff for every variable-free term t, I satisfies $\alpha[x/t]$.

When I satisfies α, we write $I \models \alpha$, and $I \not\models \alpha$, when it does not. In this section, we write $(\forall)\alpha$ and $(\exists)\alpha$ to denote $\forall x_1 \ldots \forall x_n\alpha$ and $\exists x_1 \ldots \exists x_n\alpha$, respectively, where $x_1, \ldots, x_n$ are the free variables of α.

LEMMA 205. *If I is an interpretation, then:*

1. $I \models \neg(A : \mu)$ iff $I \models (A : \neg\mu)$.

2. $I \models (\exists)(\neg(A : \mu))$ iff $I \models (\exists)(\alpha : \neg(\mu))$.

THEOREM 206. *Suppose that P is an ALP over $\mathcal{T}$. Let P' be the ALP obtained from P by replacing all annotated literals of the form $(\neg A : \mu)$ by $(A : \neg(\mu))$. Then I is a model of P iff I is a model of P'.*

Without loss of generality, we will assume (throughout this work) that ALPs contain no negated literals. Moreover, associated with every ALP P over $\mathcal{T}$, there is a function $\mathcal{T}_p$, from the class of Herbrand interpretations to the class of Herbrand interpretations, defined as follows:

$$\mathcal{T}_p(I)(A) =_{\mathsf{def}} \sqcup\{\mu : (A : \mu) = \beta_1 \wedge \ldots \wedge \beta_n \text{ is a ground instance of an annotated clause in } P \text{ and} I \models \beta_1 \wedge \ldots \wedge \beta_n\}.$$

THEOREM 207. *Suppose that P is an ALP over $\mathcal{T}$ (where $\mathcal{T}$ is a complete lattice under $\leq$) and that $\mathcal{T}_p$ is as above. Then I is a model of P iff $\mathcal{T}_p(I) \leq I$.*

THEOREM 208. *Suppose that P is an ALP over $\mathcal{T}$ as above. Then $\mathcal{T}_p$ is monotonic; that is, $I_1 \leq I_2$ entails $\mathcal{T}_p(I_1) \leq \mathcal{T}_p(I_2)$.*

The monotonicity of $\mathcal{T}_p$ guarantees, by the Tarski-Knaster theorem, that $\mathcal{T}_p$ has a least fixed point that coincides with the least pre-fixed point of $\mathcal{T}_p$. (Here, I is a pre-fixed point of $\mathcal{T}_p$ iff $\mathcal{T}(I) \leq I$.)

THEOREM 209. *P has a least model that is identical with the least fixed point of $\mathcal{T}_p$.*

Given that $\mathcal{T}$ is a complete lattice, it possesses a least element and a greatest element, which we denote, respectively, by $\perp$ and $\top$. Furthermore, associated with every $\mathcal{T}$, there are two distinguished interpretations, denoted by $\triangle$ and ∇, that assign the truth-values $\perp$ and $\top$, respectively, to the elements $A \in Bp$, where Bp is the Herbrand base of P.

DEFINITION 210. If P is an ALP over $\mathcal{T}$, then the *upward iteration* of $\mathcal{T}_p$ is defined, for all ordinals λ, as

$$\begin{cases} \mathcal{T}_p \uparrow 0 =_{\mathsf{def}} \triangle \\ \mathcal{T}_p \uparrow \lambda =_{\mathsf{def}} \sqcup_{\alpha<\lambda} \mathcal{T}_p(\mathcal{T}_p \uparrow \alpha). \end{cases}$$

THEOREM 211. *$\mathcal{T}_p \uparrow \omega$ is identical to the least fixed point of $\mathcal{T}_p$.*

DEFINITION 212. A model I of the ALP P over $\mathcal{T}$ is *supported* iff $I(A) = \sqcup\{\mu : (A : \mu) \Leftarrow (\beta_1 : \mu_1) \wedge \ldots \wedge (\beta_n : \mu_n)$ is a ground instance of an annotated clause in P and $I \models (\beta_1 : \mu_1) \wedge \ldots \wedge (\beta_n : \mu_n)\}$.

THEOREM 213. *I is a fixed point of $\mathcal{T}_p$ iff I is a supported model of P.*

DEFINITION 214. Let λ be an ordinal. The *downward iteration* of $\mathcal{T}_p$ is defined as follows:

$$\begin{cases} \mathcal{T}_p \downarrow 0 =_{\mathsf{def}} \nabla \\ \mathcal{T}_p \downarrow \lambda =_{\mathsf{def}} \sqcap_{\alpha<\lambda} \mathcal{T}_p(\mathcal{T}_p \uparrow \alpha) \end{cases}$$

DEFINITION 215. The ALP is *canonical* iff $\mathcal{T}_p \downarrow \omega$ is a fixed point of $\mathcal{T}_p$.

DEFINITION 216. Suppose C_1 and C_2 are the annotated clauses given below:

$$\begin{cases} (A^1 : \mu^1) \Leftarrow (B_1^1 : \rho_1^1) \wedge \ldots \wedge (B_n^1 : \rho_n^1) \\ (A^2 : \mu^2) \Leftarrow (B_1^2 : \psi_1^2) \wedge \ldots \wedge (B_m^2 : \psi_m^2). \end{cases}$$

C_1 is *semantically equivalent* to C_2 iff there is a substitution Θ such that $A^1\Theta = A^2$ and

$$\{(B_1^1\Theta : \rho_1^1), \ldots, (B_n^1\Theta : \rho_n^1)\} = \{(B_1^2 : \psi_1^2), \ldots, (B_m^2 : \psi_m^2)\}.$$

DEFINITION 217. Suppose that C_1 and C_2 do not share common variables. An ALP P is *closed* iff for every pair C_1, C_2 of annotated clauses in P satisfying the previous condition, if C_1, C_2 are of the form

$$\begin{cases} (A^1 : \mu^1) \Leftarrow (B_1^1 : \rho_1^1) \wedge \ldots \wedge (B_n^1 : \rho_n^1) \\ (A^2 : \mu^2) \Leftarrow (B_1^2 : \psi_1^2) \wedge \ldots \wedge (B_m^2 : \psi_m^2), \end{cases}$$

such that A^1, A^2 are unifiable via the mgu Θ, and μ^1 and μ^2 are incomparable (i.e., $\mu^1 \nleq \mu^2$ and $\mu^2 \nleq \mu^1$), it is the case that

$$((A^1 : \sqcup\{\mu^1, \mu^2\}) \Leftarrow (B_1^1 : \rho_1^1) \wedge \ldots \wedge (B_n^1 : \rho_n^1) \wedge (B_1^2 : \psi_1^2) \wedge \ldots \wedge (B_m^2 : \psi_m^2))\Theta$$

is semantically equivalent to some annotated clause in P. The *closure* of an ALP P, denoted $CL(P)$, is the closed ALP obtained by repeatedly adding to P all clauses C obtained from annotated clauses C_1, C_2 whose heads are unifiable and whose heads's annotations are incomparable.

Every ALP P can be extended to a closed ALP $CL(P)$ by adding a finite number of new annotated clauses, as is easy to show. Moreover, we have:

THEOREM 218. *Suppose P is an ALP over a complete lattice $\mathcal{T}$ of truth-values. Then:*

(1) $\mathcal{T}_p = \mathcal{T}_{CL(P)}$

(2) *Hence, P and $CL(P)$ have the same models, i.e., they are logically equivalent.*

(3)*If I is a supported model of $CL(P)$, and A is a variable-free atom such that $I(A) = \lambda \neq \perp$, then there is a single annotated clause in P having a ground instance of the form*

$$(A : \mu) \Leftarrow (B_1 : \psi_1) \wedge \ldots \wedge (B_n : \psi_n)$$

such that $\lambda \leq \mu$ and $I \models (B_1 : \psi_1) \wedge \ldots \wedge (B_n : \psi_n)$.

DEFINITION 219. P is canonical iff $\mathcal{T}_p \downarrow \omega$ is the greatest fixed point of $\mathcal{T}_p$.

THEOREM 220. *If P is canonical, then $\mathcal{T}_p \downarrow \omega$ is the greatest supported model of P.*

Executing queries

Important applications of the formal framework above are found when we analyze the interactions between an user who ask queries to a knowledge base. To discuss cases of this type, we assume that all ALPs are closed, which suffices for this purpose (see [da Costa and Subrahmanian, 1989]). In this kind of application, the knowledge base is expressed as an ALP over a suitably chosen complete lattice ($\mathcal{T}, \leq$) of truth-values. In this framework, experts may use expressions such as *p is likely to be false* or, more precisely, *p is false with 90 % certainty* (see [da Costa and Subrahmanian, 1989], where examples are given).

Other applications of annotated logics will be mentioned in the next section.

6 DEVELOPMENTS IN PARACONSISTENT LOGIC

6.1 Some carried out developments

Since the 1960s, when paraconsistent logic was established as a logic *stricto sensu*,[16] several developments have been made, most of them in connection with the C-systems.[17]

To give an idea (although not a complete one) of some of these developments,[18] let us recall *some* of the most important events. In 1969, M. Fidel proved the decidability of C_n, $1 \leq n \leq \omega$, by algebraic methods (see [Fidel, 1977], and also [D'Ottaviano, 1990]). Another decision method for C_n, $1 \leq n \leq \omega$, was presented by D. Marconi in 1980 using semantic tableaux (see [Marconi, 1980]). In 1987, W. Carnielli, after systematizing finite many-valued logics through tableaux, approached C_1 by this method, and also showed that C_1 is decidable (see [Carnielli, 1987]).

In 1970s, da Costa, Alves, Loparić and Arruda studied the semantic counterpart of the calculi C_n, $1 \leq n \leq \omega$, later extended to the calculi $C_n^\star$ and $\mathcal{D}_n$, $1 \leq n \leq \omega$ (see [Arruda and da Costa, 1977]). Da Costa and Alves' work on the issue has been mentioned already. Although the subject has not been discussed here, new hierarchies of calculi, constructed between C_n and C_{n+1}, $n \geq 0$, have also been introduced. Their semantics was studied by Alves (see [Alves, 1976]). We noted above that Alves has proved the decidability of da Costa's propositional systems by the method of quasi-matrices. Later, Loparić presented a two-valued semantics and a decision method for C_ω (see [Loparić, 1977], [Loparić, 1978], and [Loparić and Alves, 1980]).

Arruda and da Costa have also axiomatized some paraconsistent systems that are also relevant logics (see [Arruda and da Costa, 1984]). In their systems, $\alpha \wedge (\alpha \rightarrow \beta) \rightarrow \beta$, the rules of contraction, namely, $\alpha \rightarrow (\alpha \rightarrow \beta)/\alpha \rightarrow \beta$ and $(\alpha \rightarrow (\alpha \rightarrow \beta)) \rightarrow (\alpha \rightarrow \beta)$, and the deduction theorem do not hold. These systems (called P and $P^\star$), as well as their quantificational counterparts, are not finitely trivializable, are not decidable by finite matrices, but can be extended to modal and tense logics, as these authors have shown.

---

[16] That is, with the development of (at least) first-order predicate calculus. This turning point is acknowledged in general (see, for instance, [Tuziak, 1993]).

[17] Further details on these developments and references to the authors who have contributed to the subject can be found in [Arruda, 1980], [Arruda, 1989], [da Costa and Marconi, 1989], and [D'Ottaviano, 1990].

[18] For further information, see the references just mentioned.

Routley and Loparić studied the semantic aspects of P, as well as some of its 'dialectical' extensions (see [Routley and Loparić, 1980]). Arruda and da Costa have also developed the systems J_n, $1 \leq n \leq 5$ (at the propositional and predicate levels), in which modus ponens is not valid (see [Arruda and da Costa, 1970]). These systems have been studied also by M. Bunder (see [Bunder, 1983]).

Based on Jaśkowski's ideas, D'Ottaviano and da Costa introduced a three-valued propositional logic J_3 with two designated values, which turn out to be paraconsistent (see [D'Ottaviano and da Costa, 1970]). Furthermore, D'Ottaviano presented an axiomatization of J_3 (see [D'Ottaviano, 1982] and [D'Ottaviano, 1985a]), and studied the connections between this calculus and several others, such as intuitionistic logic and Łukasiewicz's three-valued logic. The system J_3 was further studied by D'Ottaviano herself in other papers, where she introduced the concept of J_3-theories, and adapted and proved to the latter several results of model theory (see also [D'Ottaviano, 1985b]).

The algebraization of Jaśkowski's logic was studied by Kotas, following the axiomatic treatment given by da Costa, Dubikajtis and Kotas himself. In particular, Kotas proved that the system is not decidable by finite matrices (see [Kotas, 1975]). Other studies related to Jaśkowski's logics have been developed in [Furmanowski, 1975], [Kotas and da Costa, 1977], [Kotas and da Costa, 1978], [Kotas and da Costa, 1979], and [Marconi, 1979]. Pinter's system that deals with 'inherent ambiguity' is a slight modification of a Jaśkowski logic (see [Pinter, 1980]).

One of the main problems with regard to the algebraic study of the C-logics is that the only congruence relation in C_1 is the identity relation, as shown by C. Mortensen (see [Mortensen, 1980]). Despite this fact, some algebraic approaches to the C-systems, or their extensions, were presented in [da Costa, 1966a], [da Costa, 1966b], [da Costa and Sette, 1969], [Sette, 1971], [Carnielli and Alcântara, 1984], and [Lewin *et al.*, 1990], as well as in some of Béziau's works referred to below.

The fact that the C-systems do not enable substitutivity by equivalents is discussed by I. Urbas (see [Urbas, 1989]), who attempted to remedy this situation by extending these systems with the addition of new rules. He then proved that these extensions do not lead to systems distinct from classical logic, concluding that new hierarchies should be constructed, where adequate equivalence relations can be formulated. The same problem regarding the impossibility of defining such equivalences in da Costa's systems is discussed by Peña in [Peña, 1991, pp. 284ff].

Important contributions to PL have been made by J. -Y. Béziau in a series of works. He started studying PL in the late 1980s, formulating a semantics for C_1. He also axiomatized a sequent system for C_1. (This was a problem that A. Raggio tried to solve in 1960s; see [D'Ottaviano, 1990].) Béziau also extended C_1 to a stronger system, C_1+, by replacing the axioms $\alpha^o \wedge \beta^o \rightarrow (\alpha \odot \beta)^o$, where $\odot \in \{\wedge, \vee, \rightarrow\}$, by $\alpha^o \vee \beta^o \rightarrow (\alpha \odot \beta)^o$. As a result, he obtained more theorems as well as some of De Morgan laws. He then studied a non-truth functional semantics for the stronger system (see [Béziau, 1990]). Béziau also investigated a general theory of negation (see [Béziau, 1995a]), and argued that both classical first-order logic and the modal system S5 can be viewed as paraconsistent systems (see [Béziau, 2002a]). These last results led him to study a new theory of opposition, where a polyhedron replaces the traditional square of opposition (see [Béziau, 2003]).

The extension of first-order paraconsistent logic to set theory has already been discussed in section 3. This work was initiated by da Costa (see [da Costa, 1963a]), and continued by Arruda (see [Arruda, 1964]). (For additional references and historical details, see [D'Ottaviano, 1990] and [da Costa *et al.*, 1998a].) Higher-order paraconsistent logics corresponding to the C-systems have been presented in [Alves and Moura, 1978], [da Costa and Alcântara, 1982], and [da Costa and Alcântara, 1986]. Semantics for paraconsistent systems containing descriptions and Hilbert's ϵ-symbol have been discussed, respectively, by Abar (see [Abar, 1985]) and Yamashita (see [Yamashita, 1985]). These authors have also shown how to develop a general theory of v.b.t.o. in the calculi $C_n^=$.

These developments are, of course, only part of what has been done in paraconsistent logic. Despite the huge amount of work that has been produced in the area (which we cannot, of course, even begin to summarize here), we still do not have a comprehensive introductory book on the subject. (A first introductory book on PL is Grana's [Grana, 1983]; see also [Grana, 1990a; Grana, 1990b; Grana, 1990c; Grana, 2001] for some of his ideas on the subject.) We will now briefly comment on some other lines of research.

6.2 *A taxonomy of C-systems*

Recently, the C-logics have been studied from a different perspective. In 'A Taxonomy of C-systems', Carnielli and Marcos presented an elaborate study of the foundations of paraconsistent logics (see [Carnielli and Marcos, 2002]). In particular, the authors show how several classes of paraconsistent logics can be distinguished from the point of view of general abstract logic. They also put forward a new discriminating account of several logical principles, such as the principle of noncontradiction and various forms of pseudo-Scotus (also known as the principle of explosion).

The logics of formal inconsistency (LFIs) are then introduced as a large class of paraconsistent logics in which the concepts of consistency and inconsistency can be internalized. This allows Carnielli and Marcos to present a novel account of the notion of consistency, and distinguish formally between the notions of contradictoriness and inconsistency. While studying the general features of an important subclass of LFIs — namely, the C-systems and the dC-systems — they show that most paraconsistent logics can be seen as C-systems (C-logics), and explore their properties and shortcomings.

Developing further the work done in [Carnielli and Marcos, 2002], Carnielli, Coniglio and Marcos improved some results and emphasized semantic and proof-theoretic aspects of the logics of formal inconsistency (see [Carnielli *et al.*, 2006]). The main LFI and one of its primary subclasses, the C-systems, were also surveyed. A striking feature of LFIs is their ability to encode classical logic, in the sense that LFIs are able to reproduce classical reasoning, despite being subsystems of classical logic. The dC-systems, a particular subclass of the C-systems, are carefully discussed; particular cases include da Costa's C_n, and Jaśkowski's D2.

By adding new axioms to the dC-systems, the authors show that it is possible to introduce a large family of logics by controlling the propagation of inconsistency. Exploring this possibility allows the definition of thousands of new logics. The paper also emphasizes the semantic meaning of LFIs, discussing in detail both the valuation semantics for LFIs and the possible-translation semantics. Another kind of semantics, the society semantics, is also addressed. Furthermore, modal extensions of LFIs and their Kripke semantics are discussed, as well as some first-order LFIs. An entire subsection of [Carnielli *et al.*, 2006] is dedicated to proof systems, especially to the tableaux method; indeed, tableaux proof systems for several LFIs are provided. Several other issues, including the difficulties of algebraizing LFIs, are also studied in both papers.

A further development of this approach is concerned with modal LFIs. Given the epistemic version of the normal logic KT, if we add to it the so-called verificationist thesis, 'A implies that it is possible to know A', it follows that 'A implies that KA'. This yields the collapse of the knowledge operator. This problem is known as the knowability paradox, and it threatens any normal modal logic that expands KT by the alethic thesis. Modal LFIs are useful in the task of avoiding the knowability paradox, because it is possible to add to the paraconsistent version of the modal logic KT the verificationist thesis without losing the modality, as shown in [Carnielli *et al.*, to appear]. In this sense, modal LFIs are able to avoid the knowability paradox while simultaneously recovering classical reasoning with such logics.

6.3 Other directions

Several directions, other than those suggested by the development of da Costa's C-logics, were also proposed to deal with inconsistencies. For instance, D. Batens initially studied paraconsistent systems related to dialectics (see [Batens, 1980]). Later, he considered the case of inconsistency-adaptive logics, i.e. logics in which abnormalities are inconsistencies, and presented two first-order inconsistency-adaptive logics, both from the proof-theoretic and the model-theoretic points of view (see [Batens, 1999] and [Batens, 2000]).

In 1979, Lorenzo Peña introduced a new kind of PL that resembled fuzzy logic in certain respects (see [Peña, 1979]; see also [Peña, 1991; Peña, 1993], where Peña's ideas are developed further).

Paraconsistent logic has also progressed in Australia and New Zealand, in part due to the effords of R. Routley, motivated by the connections with relevant logics. In these logics, certain schemas and rules from classical logic, such as $\alpha \to (\neg\alpha \to \beta)$ and $\alpha, \neg\alpha \vdash \beta$, are not valid (see [Anderson and Belnap, 1975], [Routley and Meyer, 1976], and [Routley, 1979]). As a result, these systems can also be used as the basis for inconsistent but nontrivial theories. Marconi gave a precise formulation of the relation between paraconsistent logic and relevant logic (see [Marconi, 1979]). From our point of view, most relevant systems are paraconsistent. Additional contributions from Australia can be found in the works of Mortensen [Mortensen, 1980], Bunder [Bunder, 1980; Bunder, 1983], and Priest [Priest, 1979; Priest *et al.*, 1989; Priest, 1987] (see also [Priest and Tanaka, 2000] for Priest's view on paraconsistent logics).

Another domain to which paraconsistent logic can be applied is dialectics. From a historical perspective, dialectics may implicitly assume some system of paraconsistent logic, since most dialectical views maintain that there are 'real contradictions', i.e., that the actual world is contradictory. (For a basic antology on this subject, see [Marconi, 1979].)

We mention here the researches of da Costa and R. G. Wolf on the underlying logic of dialectics — dialectics conceived according to the interpretation of McGill and Parry (see [McGill and Parry, 1948]). The resulting systems, involving propositional and predicate levels, have formal features analogous to the C-logics. But they are strictly stronger than the latter (for details, see [da Costa and Wolf, 1980] and [da Costa and Wolf, 1985]).

In [Akama, 2002], S. Akama reappraises D. Nelson's work on inconsistent systems from the 1950s (see [Nelson, 1949; Nelson, 1959]). Taking Nelson as a forerunner of PL, Akama presents an interesting study on Nelson's 'constructive' PL.[19] Additional developments of PL, more closely related to applications, will be discussed in the next section.

7 APPLICATIONS

7.1 Technology

As we discussed above, annotated logics were originally introduced by Blair and Subrahmanian in the 1980s. These logics have been developed further and applied to several fields, such as robot control [Nakamatsu *et al.*, 2002], air traffic control [Nakamatsu *et al.*, to appear], control systems for autonomous machines, defeasible deontic reasoning [Nakamatsu *et al.*, 2001], information systems [Akama and Abe, 2001] and medicine. Here, we cannot do justice to all applications that have been devised in recent years. But let us give a brief sketch of some of the recent developments in the area.

A programming language, called PARALOG, was implemented in [Ávila and Abe, 1997]. It is a paraconsistent version of PROLOG, and it has been used to construct several computational systems for planing data, vision systems, and to represent inconsistencies (see [Ávila, 1996]).

Furthermore, digital circuits, inspired in annotated logics, were introduced in [Abe and da Silva Filho, 1998] to accommodate incompatible signals. The authors suggest that this device may be useful to develop more general electric circuits, as well as to articulate applications in logistic and to decision procedures. Hardware devices are also under construction, motivated by paraconsistent ideas. The so-called *para-analyser* enables scientists to handle uncertainty, inconsistencies and paracompleteness (see [Abe and da Silva Filho, 2003]). Several other related devices have also been constructed, leading to the first 'paraconsistent robot': Emmy (see [Silva Filho and Abe, 2001a] and [Silva Filho and Abe, 2001b]).

Interesting applications are being developed in medicine: in the recognition of cancer cells, in Alzheimer illness, and in disfunctions of speech. These applications are new, and they suggest that applied paraconsistent logics may have an extraordinary role. Furthermore, in engineering, nonmonotonic and defeasible forms of reasoning have been represented in terms of paraconsistent logics, leading to the development of softwares that are being used in traffic control — of trains, aircrafts, and cars (see [Nakamatsu *et al.*, 1999] and [Nakamatsu *et al.*, 2000]). The hardware counterpart has also been developed in the form of a computer chip (see [Nakamatsu *et al.*, 2003]).

---

[19] Akama claims that 'da Costa did not appear to be familiar with Nelson's system' (see [Akama, 2002]). However, Nelson's paper [Nelson, 1959] is listed in the references of da Costa's seminal work [da Costa, 1963a].

7.2 Informatics

Several types of applications of paraconsistent logics in informatics have been devised. In this section, we will discuss some of them.

Epistemic inconsistencies in artificial intelligence

Paraconsistent logic has been used in the construction of models of 'real life reasoning' in artificial intelligence. This is part of the general problem of 'practical reasoning', and is treated under the label of the 'logic of appearance'. In all of these areas, we find contributions by T. Pequeno, A. Buchsbaum, A. T. Martins, and their collaborators (see, e.g., [Buchsbaum *et al.*, 2004], [Martins and Pequeno, 1996], [Pequeno and Buchsbaum, 1995], [Correa *et al.*, 1993], and [Pequeno, 1990]). For example, they introduce the notion of epistemic inconsistency, which refers to the existence of contradictory views about the same situation (see [Pequeno, 1991]). These contradictions reflect the incompleteness (or vagueness) of our knowledge about the domain under study. In particular, the association of this phenomenon with nonmonotonic reasoning is explored. The idea is to devise a logical system and the corresponding semantics that make these notions precise and enable reasoning on these inconsistent views without triviality (see [Pequeno, 1990]).

In [Buchsbaum and Pequeno, 1993], it is presented a proof method for automation of reasoning in paraconsistent logic, using da Costa's calculus $C_1^\star$. The method is analytical, and it employs a specially designed tableaux system. In fact, the authors present two tableaux systems. The first, with a small number of rules, is used to prove the soundness and the completeness of the method. The second, which is equivalent to the first, is a system of derived rules designed to enhance computational efficiency. A prototype based on the second system was also effectively implemented.

Other paraconsistent fuzzy systems

In addition to the systems connecting paraconsistent logic and fuzzy logic mentioned earlier, Barreto and Ebecken presented some applications of PL in artificial intelligence. They explored, in particular, the use of PL in the construction of paraconsistent knowledge bases, implemented in fuzzy shells (see [Barreto, 2004], [Barreto and Ebecken, 1998] and [Barreto and Ebecken, 2001]). It was shown that it is possible to build paraconsistent knowledge bases in the Mathlab Fuzzy Logic Toolbox, and that fuzzy shells are inconsistency-tolerant. Moreover, it has also been shown that the 'defuzzification method', that plays an important role in paraconsistent fuzzy systems, must be chosen according to the nature of the knowledge base in question (see [Barreto and Ebecken, 2002]). In fact, the authors describe a particular defuzzification method in the interpretation of a paraconsistent knowledge base.

These works provide only partial results for the handling of inconsistency in artificial intelligence (AI), and the authors insist on the need for investigating the problems of coherence and normalization of inconsistent and paraconsistent knowledge bases, including their semantic aspects. In [Barreto, 2002], the relationship between paraconsistent knowledge bases and possibilistic logic is explored. Using possibilistic logic, it is possible to accommodate inconsistent information by assigning to each piece of knowledge a certain degree of paraconsistency, and by tracking the use of such pieces in the derivation of new conclusions. This proposal seems to solve partially the problem of normalization and the problem of knowing whether paraconsistent propositions in the knowledge base affect the conclusions of the deductive processes where they are used. The degree of paraconsistency and the tracking system help to indicate to what extent the conclusions are thus affected.

Furthermore, in [Barreto, 2004], the proof of the consistency between possibilistic resolution and Zadeh's approximate reasoning theory is outlined. This is an interesting result for AI, since it provides a theoretical base for the improvement of an efficient methodology for reasoning in the presence of inconsistency. An interesting aspect of this methodology is that it provides a possible model of human knowledge, where qualitative analysis is needed. Examples of applications include medical diagnosis as well as juridical and business decisions.

The matrix connection method and paraconsistency

As is well known, W. Bibel's matrix connection method provides an alternative procedure for theorem proving other than the usual resolution technique (see [Bibel, 1982] and [Bibel, 1993]). This method was adapted and implemented in the particular case of an annotated propositional paraconsistent logic by Kaestner, Krause, Musicante and Nobre (see [Kaestner and Krause, 2996] and [Krause *et al.*, 2001]). We will give here the main ideas of this approach, and highlight some connections between annotated logic and computation.

We will be working, once again, in a propositional language with standard connectives and other symbols. Let A be a non-empty, finite, ordered set of propositional symbols. To each element $a \in A$, we associate a non-empty, finite lattice $\mathcal{T}_a$. The elements of $\mathcal{T}_a$ are called *annotated constants*, and denoted by μ, ν etc.

DEFINITION 221. A *ground literal* is a triple (a, μ, p), where $a \in A$, $\mu \in \mathcal{T}_a$, and $p \in \{0, 1\}$. p is the *polarity* of the literal. We will write either $\sim L$ or $\sim L_\mu$ to denote the literal $(a, \mu, 1)$, while either L or L_μ denotes the literal $(a, \mu, 0)$. In general, literals will be denoted by K, L, M etc.

Let R be an alphabet of *occurrences* or *positions*. The elements of R are denoted by r.

DEFINITION 222. By induction, we define the concept of *propositional matrices* over (A, R), denoted by the letters D, E, F. Similarly, we define their *size* $\sigma(F)$, their *positions* $\Omega(F) \subset R$, and the *depth* $\delta(r)$ of r in the matrix F, for any $r \in \Omega(F)$:

a) *For any literal* L, *and for any* $r \in R$, *the pair* $(L, r) = L^r$ *is a matrix with* $\sigma(L^r) = 0$, $\Omega(L^r) = \{r\}$, *and* $\delta(L^r) = 0$.

b) *If* $F_1, \ldots, F_n$, $n \geq 0$, *are matrices such that* $\Omega(F_i) \cap \Omega(F_j) = \emptyset$, *for* $i \neq j$ *and* $1 \leq i, j \leq n$, *then the set* $F = \{F_1, \ldots, F_n\}$ *is a matrix where:*

(i) $\sigma(\emptyset) = 0$, *for* $n = 0$, *and* $\sigma(F) = 1 + \sum_{i=1}^{n} \sigma(F_i)$, *for* $n > 0$;

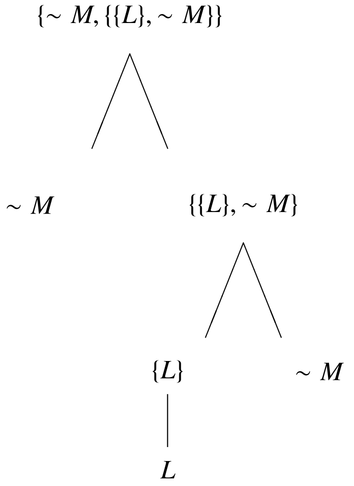


Figure 2. Representation of a matrix by a tree

(ii) $\Omega(F) = \Omega(F_1) \cup \ldots \cup \Omega(F_n)$;

(iii) $\delta(r) = m + 1$, *for any* $r \in \Omega(F_i)$, $1 \leq i \leq n$, *where* m *is the depth of* r *in* F_i.

According to this definition, the atomic parts of the matrices are ground literals, and in general, a matrix is a nested set of occurrences of literals.

EXAMPLE 223. Let us consider $A = (a, b, c, d)$, associated with the elements of the lattice FOUR (see figure 5.3), and let $R = \{0, 1, 2, 3\}$ be an alphabet of positions. Then $L = (a, \perp, 0)$ and $\sim M = (c, f, 1)$ are ground literals, while $\{\{L^0\}, \{\sim M^1\}\}$ and $\{\sim M^0, \{\{L^1\}, \sim M^2\}\}$ are matrices over (A, R).

A matrix can also be viewed as a *tree*, where some leaves are associated with literals. Figure 2 presents the tree corresponding to the second matrix of the example above.

DEFINITION 224. Let F be a matrix and $l, m \in \{0, 1\}$. The set of *formulas* $\tilde{F}$, represented by F with respect to (l, m), is inductively defined as follows:

(i) *If* F *is a literal* $F = L^r$ *and* $l = 0$, *then* $\tilde{F} = L$.

(ii) *If* F *is a literal* $F = L^r$ *and* $l = 1$, *then* $\tilde{F} = \sim L$.

(iii) *If* $F = \{F_1, \ldots, F_n\}$, $n \geq 0$, *and if* $m = 1$, *then* $\tilde{F} = \wedge(\tilde{F}_1, \ldots, \tilde{F}_n)$, *where the* $\tilde{F}_i$ *are formulas represented by* F_i *with respect to* $(l, 0)$, $i = 1, \ldots, n$.

(iv) *If* $F = \{F_1, \ldots, F_n\}$, $n \geq 0$, *and if* $m = 0$, *then* $\tilde{F} = \vee(\tilde{F}_1, \ldots, \tilde{F}_n)$, *where* $\tilde{F}_i$ *are formulas represented by* F_i *with respect to* $(l, 1)$, $i = 1, \ldots, n$.

DEFINITION 225. A formula $\tilde{F}$ is *positively represented* by a matrix F if it is represented by F with respect to $l = m = 0$. $\tilde{F}$ is *negatively represented* if it is represented by F with respect to $l = m = 1$. A *propositional formula* is any formula represented by some matrix. Formulas are also denoted by D, E, F, but this should not bring any confusion with the notation above.

We use the following notation:

1. If $n = 0$, $\wedge(F_1, \ldots, F_n)$ is abbreviated by $\mathbf{T}$, while $\vee(F_1, \ldots, F_n)$ is abbreviated by $\mathbf{F}$.

2. If $n = 1$, $\wedge(F)$ and $\vee(F)$ are both abbreviated by F.

3. If $n \geq 2$, then $\wedge(F_1, \ldots, F_n)$ is a *conjunction*, while $\vee(F_1, \ldots, F_n)$ is a *disjunction*.

4. For any literal L, the formula $\neg^k L$ is called a *hyper-literal*. If L is a hyper-literal, then $\neg^k L = (a, \neg^k(\mu), p)$, where $\neg : \mathcal{T}_a \longrightarrow \mathcal{T}_a$ denotes some fixed function (that gives the meaning to the negation), and k is a multiplicity factor (a natural number).

5. If $F = \wedge(F_1, \ldots, F_n)$ and $n \geq 0$, then $\sim F = \vee(\sim F_1, \ldots, \sim F_n)$.

6. If $F = \vee(F_1, \ldots, F_n)$ and $n \geq 0$, then $\sim F = \wedge(\sim F_1, \ldots, \sim F_n)$.

7. For any formula F, $\sim\sim F = F$.

8. Any formula $\sim F \vee G$ is abbreviated by $F \to G$.

9. Any formula $(F \to G) \wedge (G \to F)$ may be written as $F \leftrightarrow G$.

10. We also adopt the convention that the order of precedence decreases in the sequence $\sim, \wedge, \vee, \to, \leftrightarrow$. Parentheses, which are redundant given this convention, may be deleted.

Note that if F and G are formulas, then $\neg(F \to G)$ and $\neg \sim F$, for example, are not formulas.

According to the conventions above, every well-formed formula (defined in the standard way) determines a unique matrix. However, a matrix may represent more than one formula.

EXAMPLE 226. Let $F = \{K, \{L, \sim M\}, \{\}\}$ be a matrix. The tree in figure 3 is the positive representation of F.

EXAMPLE 227. The formulas $K \wedge L \to M$, $K \to \sim L \vee M$, $L \wedge K \to M$ are all represented by $\{\sim K^1, \sim L^2, M^3\}$.

The results presented in [Bibel, 1982], chapter 2, remain applicable here, such as the following:

- If a formula $\tilde{F}$ is positively represented by a matrix F, then $\sim \tilde{F}$ is negatively represented by F.

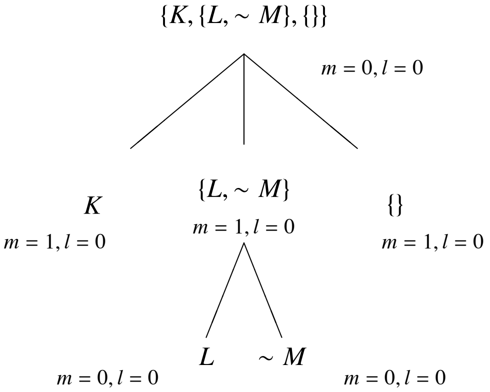


Figure 3. Positive representation of $F = \{K, \{L, \sim M\}, \{\}\}$

- If two formulas $\tilde{F}_1$ and $\tilde{F}_2$ are positively represented by the same matrix F, then $\tilde{F}_1$ and $\tilde{F}_2$ are logically equivalent in the sense of annotated logics.

These results justify the use of matrices instead of formulas. The following example justifies the name *matrix*, that was first employed by Bibel (see [Bibel, 1982]) and is also used here (see also [Kaestner and Krause, 2996]).

EXAMPLE 228. Let $\tilde{F}$ be the formula

$$(K \wedge \sim L \rightarrow \sim N) \wedge M \wedge \neg L \rightarrow (\sim N \wedge \sim K)$$

where K, L, M and N are literals. If we put $\tilde{F}$ in the disjunctive normal form (as usual), we have:

$$(K \wedge \sim L \wedge N) \vee \sim M \vee \sim \neg L \vee (\sim N \wedge \sim K)$$

This formula can be represented by a bi-dimensional arrangement, where the literals, placed in a fixed column, are connected by '$\wedge$', while the columns are connected by '$\vee$', as follows:

$$F = \begin{bmatrix} K & & & \sim N \\ \sim L & \sim M & \sim \neg L & \\ N & & & \sim K \end{bmatrix}$$

With regard to semantics, let us first introduce the following definition:

DEFINITION 229. An *interpretation* $\mathcal{M}$ is a function that associates to each propositional symbol, an element of a chosen lattice. By denoting $\mathcal{M}(a) = \mu_a$, we can write $(a, b, c, \ldots) \mapsto (\mu_a, \mu_b, \mu_c \ldots)$.

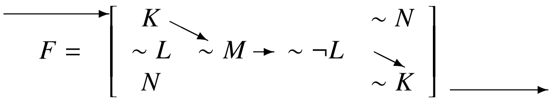


Figure 4. The path $\{K, \sim M, \sim \neg L, \sim K\}$ through the matrix F

We can now define the 'truth value' $\mathcal{M}(F)$ of a matrix F as follows:

- If F is a literal $(a, \mu, 0)$, then $\mathcal{M}(F) = \mathbf{T} = \{\emptyset\}$ iff $\mathcal{M}(a) \geq \mu$; otherwise, $\mathcal{M}(F) = \mathbf{F} = \emptyset$.
- If F is a literal $(a, \mu, 1)$, then $\mathcal{M}(F) = \mathbf{F} = \emptyset$ iff $\mathcal{M}(a) \geq \mu$; otherwise, $\mathcal{M}(F) = \mathbf{T} = \{\emptyset\}$.
- If F is a matrix $F = \{F_1, \ldots, F_n\}$, $n \geq 0$, then $\mathcal{M}(F) = \bigcup_{k=1}^{n} \mathcal{M}(F_k)$ if $m = 0$, and $\mathcal{M}(F) = \bigcap_{k=1}^{n} \mathcal{M}(F_k)$ if $m = 1$.

We write $\mathcal{M}$ `sat` $\tilde{F}$ (and also $\mathcal{M}$ `sat` F) iff $\mathcal{M}(F) = \mathbf{T}$, for a matrix F that represents $\tilde{F}$.

DEFINITION 230. A matrix F is *valid* iff $\mathcal{M}(F) = \mathbf{T}$, for every interpretation. It is called *contradictory* iff $\mathcal{M}(F) = \mathbf{F}$, for every interpretation.

Note that F is valid iff $\sim F$ is contradictory.

Paths, Connections, and Validity

DEFINITION 231. A *path* through a matrix F is a set of occurrences of literals, defined as follows:

1. *If $F = \emptyset$, then the only path through F is $\emptyset$.*
2. *If $F = L^r$, then the only path through F is the set $\{L^r\}$.*
3. *If $F = \{F_1, \ldots, F_m, F_{m+1}, \ldots, F_{m+n}\}$, $m, n \geq 0$, $m + n \geq 1$, for m literals $F_1, \ldots, F_m$ and for n matrices that are not literals $F_{m+1}, \ldots, F_{m+n}$, then for any matrix $E_i \in F_{m+i}$, and for any path p_i through E_i, $1 \leq i \leq n$, the set $\bigcup_{j=1}^{m} \{F_j\} \cup \bigcup_{i=1}^{n} p_i$ is a path through F.*

EXAMPLE 232. Let F be the matrix of the Example 228. A *path through* F is a path in the matrix from left to right, constrained to pass by the literals (that should be interpreted as 'gates'), as shown in Figure 4.

DEFINITION 233. The two literals $L = (a, \mu, p)$ and $M = (a, \nu, q)$ are *complementary* iff:

- *$p = 0$, $q = 1$ and $\nu \geq \mu$, or*
- *$p = 1$, $q = 0$ and $\mu \geq \nu$.*

DEFINITION 234. *Connections* are paths that have complementary literals as elements.

EXAMPLE 235. Let $\mathcal{T}_a$ be the lattice FOUR, $\sim K = (a, t, 1)$, $\sim L = (a, f, 1)$, $M = (a, \top, 0)$, and let the matrix $F = \{\sim K, \sim L, M\}$. The (singleton) path through F has no complementary literals. In fact, neither $\sim K$ nor $\sim L$ are complementary to M. Alternatively, it suffices to see that $\top \not\leq t$ and $\top \not\leq f$, and hence, M cannot be connected by either $\sim K$ or $\sim L$.

DEFINITION 236. We say that an ordered n-tuple $(\mu_1, \ldots, \mu_n), n > 1$, of non-comparable elements of $\mathcal{T}_a$ is a *decomposition* of μ if $\mu = \sqcup\{\mu_1, \ldots, \mu_n\}$ and there are no non-comparable elements $\{\mu'_1, \ldots, \mu'_n\}$ of $\mathcal{T}_a$ such that $\mu'_i < \mu_i$ and $\mu = \sqcup\{\mu'_1, \ldots, \mu'_n\}$. We note that this definition is efficient for small discrete lattices.

EXAMPLE 237. Consider the complete lattice FOUR. Then the only decomposition of the element $\top$ are the elements $\{t, f\}$, and hence, the matrix of the last example now is $F = \{\sim K, \sim L, \{M_1, M_2\}\}$, where $M_1 = (a, t, 0)$ and $M_2 = (a, f, 0)$. Note that K is complementary to M_1 and L is complementary to M_2. Alternatively, note that $t \leq t$ and $f \leq f$.

THEOREM 238 (Soundness and Completeness). *A matrix F is valid iff every path through F has a connection.*

Proof. Adapted from [Bibel, 1982, pp. 30-31], by using induction on the size of σ in the matrix F. ■

Checking the validity of a formula In our setting, the theorem proving technique consists in developing the following items:

a) First, we construct a matrix containing a goal and all the premises (or knowledge base).

b) We then check this matrix for validity.

Suppose that we have a set of formulas $\Gamma = \{F_1, \ldots, F_n\}$ and a query G. Then, to determine whether G is a semantic consequence of Γ, we should verify that the matrix provided by $(\bigwedge_{i=1}^{n} F_i) \to G$ is valid.

The paraconsistent component appears in this procedure given our definition of complementary literals. In this case, the existence of both a literal $L = (a, \mu, p)$ and its 'negation' $\neg L = (a, \neg(\mu), p)$ is not a sufficient condition to guarantee the existence of complementary literals in the path (see the definition 233). This illustrates the key ideas of the 'paraconsistent program' (see [da Costa, 1974]). Note that to obtain a proof of the query, it is necessary that all paths of the matrix $\sim (\bigwedge_{i=1}^{n} F_i) \vee G$ have connections. This implies the existence of complementary literals in every path.

A case study Consider the construction of a simple medical system, aimed at diagnosing three diseases K, L and M. Let us suppose that there are two different symptoms, denoted by N and O. The intended usage of the system can be described as follows:

(i) The core part of the system is the knowledge provided by a doctor (DOC_1).

(ii) Before applying this knowledge to a specific patient, say Paul, the pathologists, X-ray technicians and other professionals who conduct medical tests on Paul add the results of these tests to the knowledge base.

(iii) To use the system, we submit a goal to the program similarly to what is done in PROLOG.

Usually, in the system, the main knowledge base described above is kept in one file, while each patient's records are maintained in a separate file (or possibly as a record in a given file). The main knowledge base and this file (or its record) are merged together to form an updated knowledge base. The latter is then used in the program that diagnoses the patient's disease.

We assume that the system is written as a finite set of annotated formulas over FOUR. Suppose also that DOC_1 provided the following five rules (formulas):

(F_1) $K_t \to L_f$
(F_2) $L_t \to K_f$
(F_3) $K_t \to M_t$
(F_4) $N_t \to K_t$
(F_5) $O_t \to L_t$

Intuitively, the doctor is asserting that:

- An individual cannot have both diseases K and L. (F_1 and F_2)
- If an individual has the disease K, then he has the disease M. (F_3)
- If an individual has the symptom N, then he has the disease K. (F_4)
- If an individual has the symptom O, then he has the disease L. (F_5)

To exemplify the use and behavior of the program, we describe three situations. The first is similar to asking a query to a PROLOG program, while the other two explore the capacity of our method to handle inconsistencies:

Case 1: Suppose that the pathologist tell us that Paul was tested positively for the symptom N. And we want to know whether Paul has the disease K, but not L.

To answer this query, we must verify whether the matrix of the formula

$$F_1 \wedge F_2 \wedge F_3 \wedge F_4 \wedge F_5 \wedge N_t \to (K_t \wedge L_f)$$

is valid.

First, we transform the formula into its disjunctive normal form, and then into its matrix form. The result of this process is shown in Figure 5. Now we must check whether all paths in this matrix are complementary. Note that the first path $\{K_t, L_t, K_t, N_t, O_t, {\sim}N_t, K_t\}$ has as complementary literals N_t and ${\sim}N_t$. And it is easy to see that all paths in this matrix have complementary literals. Since the matrix is valid, we conclude that Paul has the disease K, but not L.

$$\begin{bmatrix} K_t & L_t & K_t & N_t & O_t & \sim N_t & K_t \\ \sim L_f & \sim K_f & \sim M_t & \sim K_t & \sim L_t & & L_f \end{bmatrix}$$

Figure 5. The matrix of the formula $F_1 \wedge F_2 \wedge F_3 \wedge F_4 \wedge F_5 \wedge N_t \rightarrow (K_t \wedge L_f)$

Case 2: Suppose now that the pathologist determined that Paul was tested positively for symptoms N and O. And we want to know whether Paul has both diseases K and L.

To verify this query, we must check whether the matrix of the formula

$$F_1 \wedge F_2 \wedge F_3 \wedge F_4 \wedge F_5 \wedge N_t \wedge O_t \rightarrow (K_t \wedge L_t)$$

is valid.

Following the same procedure used in the first case, it is easy to see that the matrix for this formula is valid. We conclude that Paul has both diseases K and L.

Note that this conclusion is in contradiction to what DOC_1 said: see the formulas F_1 and F_2 above. The reason for this conclusion is that the information given by the pathologist is in contradiction with that provided by the doctor. In a classical logic setting, this contradiction would make the knowledge base trivial (all formulas in the language could be derived from the knowledge base). In our case, the inconsistency is limited to the literals K and L; any atomic formula containing these literals can be proved.

Case 3: The previous example shows an inconsistent knowledge base, used to derive an inconsistent formula. The system, however, can handle inconsistencies in a nontrivial way. Let us exemplify this situation using the same set of symptoms as in Case 2 to determine whether Paul *does not have* the disease M. The latter condition can be written as M_f. So, the new situation is described by the following formula:

$$F_1 \wedge F_2 \wedge F_3 \wedge F_4 \wedge F_5 \wedge N_t \wedge O_t \rightarrow M_f$$

Following the evaluation procedure once more, we verify that the matrix for this formula is *not valid.*

The last two queries show that our method can deal with inconsistencies in the knowledge base without every formula becoming derivable.

Implementation Issues In [Krause *et al.*, 2001], an implementation of our method, using Standard ML Language [Milner, 1997], was described. But this process is not discussed here for reasons of space.

The ideas presented here can be used to model reasoning made by an intelligent agent that admits contradictory information. We believe that a system with this feature is more robust than the traditional ones. After all, the system keeps the inconsistency within a subset of its formulas, without affecting other parts of the knowledge base.

Inductive paraconsistent logic

Science does not use only deductions, of course. It is natural to ask then whether we can build 'nonclassical inductive logics' on a pair with nonclassical deductive logics, to develop systems especially in artificial antelligence (AI). John Pollock, who has been investigating several aspects of defeasible reasoning in AI, notes that:

> A common misconception about reasoning is that reasoning is deducing, and in good reasoning the conclusions follow logically from the premisses. It is now generally recognized both in philosophy and in AI that nondeductive reasoning is at least as common as deductive reasoning, and a reasonable epistemology must accommodate both. ([Pollock, 1987])

The literature on nondeductive forms of reasoning presents various systems of computational tools devoted to nonmonotonic and defeasible reasoning (see, e.g., [Pequeno, 1990] and [Pollock, 1987], to mention just two works). A particularly difficult task for AI experts is to improve AI systems so that the latter can 'reason' and 'make inferences' from vague propositions, as we humans usually do. These propositions are such that we cannot, in general, assign with certainty one of the two truth-values, *true* or *false*. Frequently, our (human) reasoning is performed by attributing only a certain 'degree of confidence' as to whether the propositions in question are true or not. Suppose, for instance, that we are visiting a foreign country, and our informant gives us a vague information about the location of a certain building. (Perhaps our informant is also not sure about the right geography of the city.) We 'believe' in the information with a certain degree of confidence, and 'decide' the way to be taken.

We will outline now of a way of accommodating such 'degrees of confidence' in terms of paraconsistent logic. The result is a different process of dealing with vagueness, which, we believe, may be of interest to AI researchers. Our main motivation is, of course, to handle vague information mechanically. But, in this section, we will limit ourselves to the description of a vague inductive logic that, we hope, will be useful in the mechanical treatment of inductive information. Further developments should lead to the development of expert systems based on the proposal below. (The main ideas are from [da Costa and Krause, 2002].)

First, we need to specify in what sense we are using the word 'inductive' in 'inductive reasoning'. We take as *inductive* any reasoning such that the truth of the premises does not entail necessarily the truth of the conclusion. Instead, the conclusion should be regarded as 'plausible' in light of the premises. There are ways of measuring this 'plausibility', but we need not discuss this issue here (for details, see [da Costa, 1986b]).

Our question then is: is there an 'inductive' paraconsistent logic? Positive answers were given, from different perspectives, in [Loparić and da Costa, 1986] and [da Costa and Krause, 2002]. In the latter work (which will be our focus here), an 'inductive annotated system' was developed to deal with vagueness and degrees of confidence. The idea is that, when we face situations of vague information, such as the one about locating a building in a foreign city, we should opt for the more prudent alternative, according to a 'warning rule'. This rule tells us to be cautious in attributing degrees of confidence. In particular, we should not accept conclusions with degrees of confidence smaller than those of the premises. This is motivated by a rational stance, since a significant feature of rationality is the attempt to optimize our rational degrees of confidence in the propositions we are concerned with.

The proposal is to extend the common way of using arguments by accepting that some propositions are vague. For instance, 'Peter is smart' is a vague proposition. However, we still assign a degree of confidence in the truth of these propositions. In other words, we believe 'Peter is smart' with some degree of confidence. Peter's mother may have a great confidence in such a proposition, but his teacher may be not so confident. The degree of confidence can be interpreted as the amount of confidence someone has in the truth of a proposition.

To accommodate these two concepts, namely, vagueness and degree of confidence, we use the resources of annotated logics.

Let us call $\mathcal{I}_\tau$ a propositional logic whose language has the following categories of primitive symbols: (a) a countable set of propositional letters, which stand for propositions (we use $P, Q, \ldots$ as syntactic variables for propositions); (b) the elements $\mu, \ldots, \mu_1, \ldots$ of a complete lattice τ ordered by $\leq$, which are called *the values of vagueness*; (c) the usual logical connectives ($\neg$, $\wedge$, $\vee$, $\rightarrow$), as well as (d) auxiliary symbols (parentheses).

The concept of *formula* of $\mathcal{I}_\tau$ is introduced in the following way:

(i) If P is a propositional letter and $\mu \in \tau$, then $P : \mu$ is a formula of $\mathcal{I}_\tau$ (atomic formula).

(ii) If α and β are formulas, then $\neg\alpha$, $\alpha \wedge \beta$, $\alpha \vee \beta$, $\alpha \rightarrow \beta$ are formulas.[20]

(iii) Every formula is obtained only from the two clauses above.

Intuitively speaking, $P : \mu$ means that P is true with degree of vagueness μ. Note that we are assigning degrees of vagueness to atomic formulas only, and not to formulas in general. So, expressions such as:

$$((P : \mu_1) \vee (Q : \mu_2)) : \mu \tag{12}$$

are *not* well formed in our system.

DEFINITION 239.

(i) *If P is a propositional letter and $\mu \in \tau$, then:*

(a) $\neg^0 P : \mu$ *means* $P : \mu$

(b) $\neg^1 P : \mu$ *means* $\neg(P : \mu)$

(c) $\neg^k P : \mu$ *means* $\neg(\neg^{k-1}(P : \mu))$, *where k is a natural number,* $k \neq 0$.

---

[20] $\alpha \leftrightarrow \beta$ is introduced in the standard way.

(ii) *Let* $\sim\colon \tau \longrightarrow \tau$ *be a fixed mapping.*[21] *From now on, we will write* $\sim \mu$ *instead of* $\sim (\mu)$. *If* $\mu \in \tau$, *then:*

(a) $\sim^0 \mu$ *means* μ

(b) $\sim^1 \mu$ *means* $\sim \mu$

(c) $\sim^k \mu$ *means* $\sim (\sim^{k-1} \mu)$, *where* $k \neq 0$ *is a natural number.*

Expressions such as $P : \mu$ are called *annotated atoms*, while $\neg^k(\alpha : \mu)$ are *hyper-literals* of order k ($k \geq 0$). The other formulas are called *complex*.

Semantics Let τ be the complete lattice above with least element $\perp$ and greatest element $\top$. Let $h : \mathcal{P} \longrightarrow \tau$ be a mapping, called an *interpretation* of $\mathcal{I}_\tau$, where $\mathcal{P}$ is the collection of propositional letters of $\mathcal{I}_\tau$. The image of the proposition P by the mapping h will be denoted $P : \mu$, where $\mu \in \tau$. Informally speaking, as we noted, $P : \mu$ means that P is true with degree of vagueness μ. To each interpretation h, we associate a *valuation* $v_h : \mathcal{F} \longrightarrow \{0, 1\}$, where $\mathcal{F}$ is the collection of formulas of $\mathcal{I}_\tau$ defined above. Intuitively speaking, 1 and 0 stand for 'true' and 'false', respectively.

Particular applications may demand appropriate choices of the complete lattice. Here, to accommodate the case studies mentioned above, we will be concerned with a particular finite, linearly ordered set $\tau = \{\mu_1, \ldots, \mu_4\}$ (with $\mu_1 \leq \cdots \leq \mu_4$) to express the distinct degrees of vagueness of a proposition. But, clearly, the approach presented here is quite general.

DEFINITION 240. Let h and v_h be as above. Let P be a propositional letter, and let α and β denote formulas. Then:

(i) $v_h(P : \mu) = 1$ *iff* $\mu \leq h(P)$.

(ii) $v_h(\neg^k(P : \mu)) = v_h(\neg^{k-1}(P :\sim \mu))$, *where* $k \neq 0$.

(iii) $v_h(\alpha \wedge \beta) = 1$ *iff* $v_h(\alpha) = v_h(\beta) = 1$.

(iv) $v_h(\alpha \vee \beta) = 1$ *iff* $v_h(\alpha) = 1$ *or* $v_h(\beta) = 1$.

(v) $v_h(\alpha \rightarrow \beta) = 1$ *iff either* $v_h(\alpha) = 0$ *or* $v_h(\beta) = 1$.

(vi) If α *is a complex formula, then* $v_h(\neg\alpha) = 1$ *iff* $v_h(\alpha) = 0$.

If $v_h(\alpha) = 1$, we say that v_h *satisfies* α, and that it does not satisfy α otherwise (that is, when $v_h(\alpha) = 0$). If Γ is a set of formulas, then we say that a formula α is a *semantic consequence* of (the formulas of) Γ, and write $\Gamma \models \alpha$, iff for every valuation v_h such that $v_h(\beta) = 1$ for each $\beta \in \Gamma$, $v_h(\alpha) = 1$. A formula α is *valid* iff $\emptyset \models \alpha$, and in this case we write $\models \alpha$.

---

[21] The specific definition of this mapping depends on the particular application. For instance, if we take τ to be the unit interval $[0, 1] \subseteq \Re$ and $\sim (x) =_{\mathsf{def}} 1 - x$, the introduction of 'fuzzy' ways of reasoning can be performed within the scope of annotated logics (see [da Costa *et al.*, 1991]).

As usual, we say that a valuation v_h is a *model* of a set of formulas Γ iff $v_h(\beta) = 1$, for every $\beta \in \Gamma$. In particular, v_h is a model of α iff $v_h(\alpha) = 1$. Other concepts, such as maximal nontrivial sets of formulas, are defined in the usual way.

The postulates of $\mathcal{I}_\tau$ If α, β, and γ are formulas and P is a propositional letter, then the postulates (axioms plus inference rules) of $\mathcal{I}_\tau$ are the following:[22]

(I1) All the postulates of classical positive logic.

(I2) If α and β are complex formulas, then the following is an axiom: $(\alpha \to \beta) \to ((\alpha \to \neg\beta) \to \neg\alpha)$.

(I3) If α is complex, then $\alpha \vee \neg\alpha$ is an axiom.

(I4) If α is complex and β is an arbitrary formula, then $\alpha \to (\neg\alpha \to \beta)$ is an axiom.[23]

(I5) $P :\perp$ is an axiom.[24]

(I6) If $\lambda \leq \mu$, then $P : \mu \to P : \lambda$

(I7) $\neg^k(P : \mu) \leftrightarrow \neg^{k-1}(P :\sim\mu)$, if $k \neq 0$.

(I8) Let α be an arbitrary formula. In this case, if $\alpha \to (P : \mu_i)$, $i \in I$, then $\alpha \to (P : \bigsqcup_{i \in I} \mu_i)$. If τ is a finite lattice, then this axiom may be replaced by the following one (see [da Costa *et al.*, 1991]):

$$P : \mu_1 \wedge \ldots \wedge P : \mu_n \to P : \bigsqcup_{i=1}^{n} \mu_i \tag{13}$$

The syntactic concepts of $\mathcal{I}_\tau$ are introduced in the standard way, as well as, in particular, the symbol of deduction $\vdash$ (see [da Costa *et al.*, 1991]).

We can prove the soundness and completeness of the logic $\mathcal{I}_\tau$ with respect to the semantics described in the previous section, and we will outline the result below. Let us first introduce a definition:

DEFINITION 241 (Strong Negation).

$$\neg^*\alpha =_{\mathsf{def}} \alpha \to ((\alpha \to \alpha) \wedge \neg(\alpha \to \alpha)) \tag{14}$$

It is easy to prove that $\neg^*$ has all the properties of classical negation. Hence, classical laws hold when $\neg^*$ is used instead of $\neg$ in the formulas of our system. For instance, *reductio ad absurdum*, $(\alpha \to \beta) \to ((\alpha \to \neg^*\beta) \to \neg^*\alpha)$, is a theorem of $\mathcal{I}_\tau$, and so is excluded middle, $\alpha \vee \neg^*\alpha$. Note that we can also show that if α is a complex formula, then $\neg\alpha \leftrightarrow \neg^*\alpha$ is valid (see the theorem below). However, this does not hold for formulas in general. For instance, if Q is a hyper-literal, then $\neg Q \leftrightarrow \neg^*Q$ is not valid in general (see [da Costa *et al.*, 1991]).

---

[22] These postulates are adapted from [da Costa *et al.*, 1991] and [da Costa *et al.*, 1991].

[23] In other words, classical logic holds for complex formulas. The presence of inconsistencies will be allowed at the level of atomic formulas only (see [da Costa *et al.*, 1991]).

[24] The technical reason for using this axiom is that $v_h(P :\perp) = 1$ iff $h(\alpha) \geq 0$, which is always true.

Other results are the following:

THEOREM 242.

(i) *If* $\Gamma, \alpha \vdash \beta$, *then* $\Gamma \vdash \alpha \to \beta$ *(deduction theorem).*

(ii) *If* $\Gamma \vdash \alpha$ *and* $\Gamma \vdash \alpha \to \beta$, *then* $\Gamma \vdash \beta$

(iii) $\alpha \wedge \beta \vdash \alpha$, $\alpha \wedge \beta \vdash \beta$, $\alpha, \beta \vdash \alpha \wedge \beta$

(iv) $\alpha \vdash \alpha \vee \beta, \beta \vdash \alpha \vee \beta$

(v) $\Gamma, \alpha \vdash \gamma$ *and* $\Gamma, \beta \vdash \gamma$, *then* $\Gamma, \alpha \vee \beta \vdash \gamma$ *(proof by cases)*

(vi) $\Gamma, \alpha \vdash \beta$ *and* $\Gamma, \alpha \vdash \neg^{*}\beta$, *then* $\Gamma \vdash \neg^{*}\alpha$ *(reductio ad absurdum)*

(vii) $\alpha, \neg^{*}\alpha \vdash \beta$, $\neg^{*}\neg^{*}\alpha \vdash \alpha$, $\alpha \vdash \neg^{*}\neg^{*}\alpha$

(viii) *If* α *is complex, then* $\neg^{*}\alpha \leftrightarrow \neg\alpha$.

(ix) $(\alpha : \mu_i)_{i \in I} \vdash \alpha : \bigsqcup_{i \in I} \mu_i$

(x) *If* $\Gamma \vdash \alpha$, *then* $\Gamma \models \alpha$ *(soundness theorem).*

To prove the completeness theorem, we need a few definitions and results.

DEFINITION 243.

(i) $\overline{\Gamma} =_{\mathrm{def}} \{\alpha : \Gamma \vdash \alpha\}$

(ii) Γ *is trivial iff* $\overline{\Gamma} = \mathcal{F}$, *where* $\mathcal{F}$ *is the set of formulas of* $\mathcal{I}_\tau$. *Otherwise,* Γ *is nontrivial.*

(iii) Γ *is inconsistent iff there exists* α *such that both* α *and* $\neg\alpha$ *belong to* $\overline{\Gamma}$. *Otherwise,* Γ *is consistent.*

(iv) Γ *is strongly inconsistent iff there exists* α *such that both* α *and* $\neg^{*}\alpha$ *belong to* $\overline{\Gamma}$. *Otherwise,* Γ *is strongly consistent.*

It is easy to see that Γ is strongly inconsistent iff it is trivial, and that Γ is strongly consistent iff it is nontrivial. Furthermore, for a suitable choice of τ, we can prove that there exist inconsistent but nontrivial sets of formulas, which are still not strongly inconsistent (see [da Costa and Subrahmanian, 1989]). So, the logic $\mathcal{I}_\tau$ is paraconsistent. This means that there exist interpretations h and formulas α such that $v_h(\alpha) = v_h(\neg\alpha) = 1$. But we can also prove that for a certain τ, there are formulas α and interpretations h such that $v_h(\alpha) = v_h(\neg\alpha) = 0$. So, $\mathcal{I}_\tau$ is also a paracomplete logic. These results are treated in detail in several papers listed in the References (see, e.g., [da Costa *et al.*, 1991] and [da Costa and Subrahmanian, 1989]).

LEMMA 244. *Every nontrivial set of formulas is a subset of some maximal nontrivial set of formulas.*

Proof. See [da Costa *et al.*, 1991] and [da Costa and Subrahmanian, 1989]. ∎

The completeness theorem results from the following lemma:

LEMMA 245. *If* Γ *is a maximal nontrivial set of formulas, then its characteristic function* $\chi_\Gamma : \mathcal{F} \longrightarrow \{0, 1\}$ *is a model of* Γ*; that is, this mapping is a valuation such that* $\chi_\Gamma(\beta) = 1$*, for every* $\beta \in \Gamma$.

Proof. The idea is to define, for a given interpretation h, a valuation v_h, in such a way that the rules of $\mathcal{I}_\tau$ are 'preserved'. This means that, given Γ, we may define $h : \mathcal{P} \longrightarrow \tau$ such that for every proposition $P \in \mathcal{P}$,

$$h(P) =_{\mathsf{def}} \bigsqcup_i \{\mu_i : \mu_i \in \Gamma\} \tag{15}$$

It is not difficult to prove that the valuation generated by this interpretation coincides with the characteristic function χ_Γ. ∎

As a consequence, we have the completeness theorem:

THEOREM 246. *If* $\Gamma \models \alpha$*, then* $\Gamma \vdash \alpha$.

Proof. See [da Costa and Subrahmanian, 1989] and [da Costa *et al.*, 1991]. ∎

Degrees of confidence We will sketch now a *theory of confidence*, which enables us to assign degrees of confidence to propositions, even to vague ones. Our degrees of confidence are in general only qualitative, characterized by the elements of an appropriate lattice with least and greatest elements. In abstract terms, degrees of confidence are elements of a lattice σ, and are assigned to the formulas of the language $\mathcal{I}_\tau$ when the propositional variables are interpreted as denoting vague statements (in the sense discussed above).

Let σ be a lattice with least and greatest elements, denoted respectively by $\bot'$ and $\top'$. The algebraic lattice operations are represented by $\sqcap$ and $\sqcup$, and the corresponding partial order by $\leq$. Let $\mathcal{F}$ be the set of formulas of the logic $\mathcal{I}_\tau$, and let α and β denote arbitrary annotated propositions. We then define the mapping $\mathsf{C} : \mathcal{F} \longrightarrow \sigma$, which satisfies the following postulates:

(C1) $\mathsf{C}(\alpha \wedge \neg^* \alpha) = \bot'$

(C2) $\mathsf{C}(\alpha \vee \neg^* \alpha) = \top'$

(C3) $\mathsf{C}(\bigvee_{i \in I} \alpha_i) \geq \bigsqcup_{i \in I} \mathsf{C}(\alpha_i)$, where I is finite.

(C4) $\mathsf{C}(\bigwedge_{i \in I} \alpha_i) \leq \sqcap_{i \in I} \mathsf{C}(\alpha_i)$, where I is finite.

(C5) If $\vdash \alpha \leftrightarrow \beta$, then $\mathsf{C}(\alpha) = \mathsf{C}(\beta)$.

Any mapping such as C is called a *confidence function.* Depending on the application under consideration, we can extend the postulates above. For instance, we can add the following ones: (i) if $\vdash \alpha \to \beta$, then $\mathsf{C}(\alpha) \leq \mathsf{C}(\beta)$, and (ii) $\mathsf{C}(\alpha) \sqcup \mathsf{C}(\neg^{*}\alpha) = \top\prime$.

Our proposal is to combine the concept of vagueness with that of confidence, that is, the logic $\mathcal{I}_\tau$ with the operator C. This is similar to the introduction of probability measures in a classical system of the probability calculus. Obviously, an 'algebra of confidence', as we suggest here, extends the classical case of subjective measure of probability, defined on Boolean algebras of propositions. Moreover, our procedure also encompasses Zadeh's theory of possibility (see [Zadeh, 1978] and [Dubois and Prade, 1988]).

When the degrees of confidence reduce to strict degrees of belief (that is, to subjective probability), they can be handled in the standard way, following the rules of the Bayesian probability calculus. In the more general case, it is important to strengthen our system with additional rules. In particular, we add the following warning rule (note that, below, we write $P : \mu : \lambda$ for $\mathsf{C}(P : \mu) = \lambda$):[25]

[Warning rule]

$$\frac{P : \mu_i : \lambda_i \,,\ P : \mu_j : \lambda_j}{P : \mu_i \sqcap \mu_j : \lambda_i \sqcup \lambda_j} \quad \Gamma \tag{16}$$

Γ stands for the set of contextual conditions, which provide the basis for the application of the rule. These conditions are accepted as true.

The warning rule has an intuitive appeal, and provides a tool to deal simultaneously with vagueness and degrees of confidence. Some inductive rules, such as those of induction by simple enumeration and analogy (see [da Costa, 1993]), complemented by the approach to vagueness and the degrees confidence suggested here, are also useful to reinforce the logic we are trying to build. Similarly, the methods of the theory of possibility can be useful here.

We designate the resulting system of inductive logic by $\mathbf{I}_\tau$. Our main goal is to use an appropriate $\mathbf{I}_\tau$ to the mechanization of inductive inferences, say in robotics and in the theory of expert systems. Let us outline how this can be done by considering a simple example.

Suppose an insurance company needs to provide a way of classifying people into disjoint classes by age. For instance, the company needs to distinguish young from old people to differentiate among several prices for insurance premiums. The company will run a series of interviews to determine the age of the interviewees. Suppose also that the company decides to classify people in four different categories: (C_1) 'young' (less or equal to 25 years old); (C_2) 'not so young' (less or equal to 35 years old, and more than 25); (C_3) 'not so old' (between 35 and 45 years old), and (C_4) 'old' (more than 45 years old). Furthermore, suppose that the government decided to pay for insurance services also for people who don't have birth certificates, and in some cases, who don't know precisely their age. The hard and poor conditions of life these people face may confuse the interviewers if the latter try to attribute precise ages to the interviewees.[26]

---

[25] Note also that, to state the rule, we used for simplicity $\sqcap$ and $\sqcup$ to denote the algebraic operations in both lattices τ and σ. This should not cause any confusion.

[26] This situation is not uncommon in developing and poor countries.

As a result, the information provided by the interviewers may be considered, in some cases, as not completely precise. For instance, interviewer A may report to the company his belief that John seems to be ‘not so old’ (that is, it seems to the interviewer that John has an age between 35 and 45 years). Similarly, the interviewer may report that Paul looks ‘not so young’. In other words, John and Paul were classified as satisfying, respectively, the predicates C_3 and C_2 above. But the interviewer is not completely sure about the correctness of the classification. If John and Paul have their birth certificates, in principle there would not be a problem as to how to deal with the information. After all, the interviewer’s judgment about the age of each interviewee can be double checked.

Suppose, however, that neither John nor Paul have their birth certificates. How should we to deal with this situation? Since the attribution of age cannot be done with precision, the interviewer can adopt one of the following two alternatives: (1) to classify John as ‘old’ and Paul as ‘not so young’, or (2) to classify John as ‘not so old’ and Paul as ‘young’. Several factors are involved in the choice between these two options. They include: the expertise of the interviewer (that is, her experience in the job); the interviewer’s interest in the defense of a particular policy; or an ‘external recommendation’, for instance, the company’s interest in classifying people in certain classes of ages so that the due premium prices are the highest possible (in this way, say, the company will receive more money from the government).[27]

Humans in general go around situations of this kind by accepting some (perhaps *ad hoc*) criterion. For example, they may decide what to do by adopting some rule more or less arbitrarily, or by deciding on a case by case basis. But what happens if the situation should be handled by an expert system? In this case, is it possible to keep the system with the capacity of dealing with propositions to which different ‘degrees of confidence’ are attached to? Or is it possible to follow an insurance policy wherever it takes to express the ‘confidence’ in the information provided by an interviewer?

Let τ be a linearly ordered set $\tau = \{\mu_1, \dots, \mu_4\}$ that is also a complete lattice (where $\mu_1(=\bot) \leq \dots \leq \mu_4(=\top)$). Suppose that two interviewers[28] give as inputs $P : \mu_2$ and $P : \mu_3$, and suppose that P stands for ‘John is old’. Then, according to the example we are discussing, $P : \mu_2$ means ‘John is not so young’, while $P : \mu_3$ says that ‘John is not so old’. More precisely, the first interviewer has classified John as having an age between 25 and 35 years, while the second interviewer admitted that he is between 35 and 45 years old.

To attach a certain value to John’s taxes, an expert system may follow a rule such as this: Let $C_1, \dots, C_4$ be the classes of ages in the example above. *In case of doubt as to whether someone belongs to a class C_i or C_{i+1}, classify him/her as belonging to the class C_i.* So, in the case above, John would be considered as ‘not so young’, and then his due taxes would be supposed to be smaller than if he were classified as being ‘not so old’.

The choice of $P : \mu_2$ can be interpreted as resulting from the application of the warning rule. After all, we assigned the greater degree of confidence to the vaguer proposition. In other words, recalling the convention made when we stated the warning rule, if we write $P : \mu_2 : \lambda_2$ for $\mathsf{C}(P : \mu_2) = \lambda_2$ and $P : \mu_3 : \lambda_3$ for $\mathsf{C}(P : \mu_3) = \lambda_3$, then we can say that from the 'premisses' $P : \mu_2$ and $P : \mu_3$, we arrived at the 'conclusion' $P : \mu_2$, which has greater degree of confidence. This can be expressed by saying that the 'conclusion' is $P : \mu_2 \sqcap \mu_3 : \lambda_2 \sqcup \lambda_3$. (Note that τ is a linear lattice in which $\mu_2 \sqcap \mu_3 = \mu_2$.)

---

[27] Of course, the situation could be precisely the opposite.

[28] This number can be generalized, of course.

Roughly speaking, the warning rule insists that an expert system, elaborated to accommodate vagueness and equipped with degrees of confidence, when faced with situations such as the one considered, should opt for the more prudent option. This is a perfectly rational move. After all, as noted above, rationality involves the attempt to optimize our rational degrees of confidence in the propositions under study, but with the caution of not taking conclusions with degrees of confidence smaller than those attributed to the premises.

7.3 Foundations of physics

Logic and physics

During the International Congress of Mathematicians, held in Paris in 1900, David Hilbert presented a list of 23 Problems of Mathematics that, in his view, should occupy the efforts of mathematicians in the following century. To solve one of these problems was to achieve something extraordinary in mathematics, and several Fields medals were awarded for this kind of effort. The sixth problem of Hilbert's celebrated list dealt with the axiomatization of physical theories. Hilbert proposed 'to treat in the same manner [as Hilbert himself had done with geometry], by means of axioms, those physical sciences in which mathematics plays an important part' ([Hilbert, 1902]).

In the 20th century, much was done in this direction, continuing the efforts already developed in the previous century. There have also been a parallel development of logic in the 20th century, and the development of nonclassical systems, linked to some philosophical views about science and the presentation of scientific theories, such as that one of the logical empiricists, with emphasis in logic and language. These developments forced philosophers to acknowledge that underlying the axiomatic version of a scientific theory, there are also logical postulates (often of a higher-order logic kind), which provide the grounds for the theory's deductive and mathematical counterparts. Roughly speaking, we can think as 'logical' those postulates of first-order logic (with identity), while the 'mathematical' postulates can be those of a set theory, such as Zermelo-Fraenkel. (Of course, alternative approaches can also be developed, but we will not discuss this point here.) In this framework, logic plays an important role, and if there are reasons to suspect that some logical system, other than classical logic, should be used to axiomatize a certain domain, its details must be made explicit.

For example, discussions about the relationship between logic and quantum physics tend to relate the subjects with so-called 'quantum logics'. This is a field that has its 'official' birth in Birkhoff and von Neumann's well-known paper from 1936. This is completely justified, for this fundamental work led to the development of a new field of research in logic. Today, there are various 'quantum logical systems', including some paraconsistent quantum logics (see [Dalla Chiara and Giuntini, 1989], [Dalla Chiara and Giuntini, 2000], and [Dalla Chiara and Giuntini, 2001]). However, these systems have been studied especially as pure mathematical systems, far from applications to the axiomatization of the microphysical world and from the insights of the forerunners of quantum mechanics.

Of course, the axiomatization of a given empirical theory is not always completely determinate, and the need for a logic distinct from the classical as the underlying logic of quantum theory is still an open question. The axiomatic basis of a scientific theory depends on several aspects of the theory, some more explicit than others, that are needed to take into account the theory's structure. For instance, Ludwig studies an axiomatization of quantum mechanics based on classical logic (see [Ludwig, 1990]). All such approaches, from those that use a paraconsistent logic, or some other kind of logic, to those of Ludwig, are in principle acceptable. After all, they provide different perspectives to the same domain of discourse, and such different 'perspectives' of a domain of science may demand distinct logical apparatuses. This is a philosophical stance radically different from the classical, and it is against the idea that there is only one logic — classical logic.

As we noted in the Introduction, the possibility of using nonstandard logical systems in the foundations of physics — and, more generally, in the foundations of science — does not entail that classical logic is wrong, or that quantum theory, in particular, *needs* another logic. Physicists are likely to continue to use classical, informal logic in the near future. But we should realize that other forms of logic may help us to understand better certain aspects of the quantum world that may not be so easily treated by classical means. (An interesting example is the concept of complementarity, which can be studied very naturally in a nonclassical context.) Of course, only in the future it will be decided (if at all) whether the use of nonclassical systems in science should (or should not) be preferred to the development of theories in a 'classical' framework. Clearly, any such decision will involve significant pragmatic factors.

To summarize our proposal, we do not believe that there is only one 'true logic', and we think that distinct logical systems systems, such as those provided by paraconsistent logic, are useful to approach different aspects of such wide domains of knowledge, such as quantum theory. The important point is scientists should be open to the justifiable revision of concepts (including logical ones). This point was very lucidly emphasized by Niels Bohr, who wrote:

> For describing our mental activity, we require, on the one hand, an objectively given content to be placed in opposition to a perceiving subject, while, on the other hand, as is already implied by this assertion, no sharp separation between object and subject can be maintained, since the perceiving subject also belongs to our mental content. From these considerations, it follows not only the relative meaning of every concept, or rather of every word (the meaning depending upon our arbitrary choice of viewpoint), but also that we must, in general, be prepared to accept the fact that a complete elucidation of one and the same object may require diverse points of view that defy a unique description. Indeed, strictly speaking, the conscious analysis of any concept stands in a relation of exclusion with its immediate application. The necessity of invoking a complementarity, or reciprocal, mode of description is perhaps most familiar to us from psychological problems. In opposition to

> this, the feature that characterizes the so-called exact sciences is, in general, the attempt to obtain uniqueness by avoiding all reference to the perceiving subject. This effort is found most consciously, perhaps, in the mathematical symbolism that presents for our contemplation an ideal of objectivity for whose achievement scarcely any limits are set, as long as we remain within a self-contained field of applied logic. In natural sciences, however, there can be no question of a strictly self-contained field of application of the logical principles, since we must continually expect the appearance of new facts, the inclusion of which within the boundaries of our earlier experience may require a revision of our fundamental concepts. (*ibid.*)

A case involving paraconsistency

In this section, we will discuss, as an illustration, how paraconsistent logic can be used to approach what we call 'complementarity theories'. These theories are based on Bohr's view on complementarity, a concept he introduced in his famous 'Como Lecture', in 1927, although the basic ideas go back to 1925. The consequences of his views were fundamental, particularly to the development of the Copenhagen interpretation of quantum mechanics, and they constitute, as is largely recognized in the literature, as one of the most significant contributions to the development of quantum theory. However, we are not making an exegesis of Bohr's ideas, for, as is well known, they are quite controversial.[29] Rather, we will use a particular understanding of the notion of complementarity to provide the basis for defining a general class of theories (C-theories). As will become clear, the logic of such theories is a particular paraconsistent logic, called *paraclassic*.

Roughly speaking, we say that a theory T admits a complementarity interpretation, or that T is a C-theory, if T encompasses 'true' formulas α and β that are 'mutually exclusive', in the sense that their conjunction yields a contradiction if classical logic is applied. More precisely, let $\vdash$ be the symbol of deduction of classical logic. We say that T is a C-theory if there exists γ such that if α and β are complementary, we have (in T) $\alpha, \beta \vdash \gamma \wedge \neg\gamma$. Of course, if the underlying logic of T is classical logic, since T involves complementary propositions, T is contradictory, or inconsistent.

Let us call C an axiomatized system for the classical propositional calculus (the developments presented here can be extended to quantification). The concept of deduction in C is the standard one. We use the symbol $\vdash$ to represent deductions in C. Furthermore, the formulas of C are denoted by Greek lowercase letters, while Greek uppercase letters stand for sets of formulas. The symbols $\neg$, $\rightarrow$, $\wedge$, $\vee$ and $\leftrightarrow$ have their usual meanings, and standard conventions in writing formulas will be also assumed without further comments.

DEFINITION 247. Let Γ be a set of formulas of C, and let α be a formula (of the language of C). Then we say that α is a (syntactic) P-*consequence* of Γ, and write $\Gamma \vdash_{\mathsf{P}} \alpha$, if and only if

---

[29] For references and further discussion of this particular point, see [Beller, 1992] and [da Costa and Krause, 2004].

(P1) $\alpha \in \Gamma$, or

(P2) α is a classical tautology, or

(P3) there exists a consistent (according to classical logic) subset $\Delta \subseteq \Gamma$ such that $\Delta \vdash \alpha$ (in classical logic).

We call $\vdash_{\mathsf{P}}$ the relation of P-*consequence.*

DEFINITION 248. P is the logic whose language is that of C and whose relation of consequence is that of P-consequence. This logic will be called *paraclassic.*

It is immediate that, among others, the following results can be proved:

THEOREM 249.

1. *If α is a theorem of the classical propositional calculus* C *and if* Γ *is a set of formulas, then* $\Gamma \vdash_{\mathsf{P}} \alpha$*; in particular,* $\vdash_{\mathsf{P}} \alpha$.
2. *If* Γ *is consistent* (*according to* C), *then* $\Gamma \vdash \alpha$ (*in* C) *iff* $\Gamma \vdash_{\mathsf{P}} \alpha$ (*in* P).
3. *If* $\Gamma \vdash_{\mathsf{P}} \alpha$ *and if* $\Gamma \subseteq \Delta$, *then* $\Delta \vdash_{\mathsf{P}} \alpha$. *(The notion of* P*-consequence is monotonic.)*
4. *The notion of* P*-consequence is recursive.*
5. *Since the theses of* P *(that is, valid formulas of* P*) are those of* C, P *is decidable.*

DEFINITION 250. A set of formulas Γ is P-*trivial* iff $\Gamma \vdash_{\mathsf{P}} \alpha$, for every formula α. Otherwise, Γ is P-*nontrivial.* (Similarly, we define the concept of a set of formulas being *trivial* in C.)

DEFINITION 251. A set of formulas Γ is P-*inconsistent* if there exists a formula α such that $\Gamma \vdash_{\mathsf{P}} \alpha$ and $\Gamma \vdash_{\mathsf{P}} \neg\alpha$. Otherwise, Γ is P-*consistent.*

THEOREM 252.

1. *If α is an atomic formula, then* $\Gamma = \{\alpha, \neg\alpha\}$ *is* P*-inconsistent, but* P*-nontrivial.*
2. *If the set of formulas* Γ *is* P*-trivial, then it is trivial (according to classical logic). If* Γ *is nontrivial, then it is* P*-nontrivial.*
3. *If* Γ *is* P*-inconsistent, then it is inconsistent according to classical logic. If* Γ *is consistent according to classical logic, then* Γ *is* P*-consistent.*

A semantic analysis of P, for instance a completeness theorem, can be obtained without difficulty, as indicated in [da Costa and Vernengo, 1999]. Note that the set $\{\alpha \wedge \neg\alpha\}$, where α is a propositional variable, is trivial according to classical logic, but it is not P-trivial. However, we are not suggesting that complementary propositions should be understood necessarily as pairs of contradictory sentences. This is made clear by the following definition:

DEFINITION 253 (Complementarity Theories or C_{mp}-theories). A *C-theory* is a set of formulas T of the language of C (the classical propositional calculus) closed by the relation of P-consequence, that is, $\alpha \in T$, for any α such that $T \vdash_{\mathsf{P}} \alpha$. In other words, T is a theory whose underlying logic is P. A *C_{mp}-theory* is a C-theory that satisfies additional meaning postulates (that are introduced in the particular context in which the theory is used).

Of course, the definition of a C_{mp}-theory is vague. But, in each context of use, what a C_{mp}-theory amounts to is clear. Moreover, note also that if a meaning postulate introduces restrictions in the acceptable statements of the theory, the hypotheses and axioms used in deductions must satisfy such restrictions as well. For instance, if a meaning postulate of a theory T is formulated as Heisenberg's uncertainty principle, this will impose obvious restrictions to certain statements of T.

THEOREM 254. *There exist C-theories and C_{mp}-theories that are inconsistent, although they are* P*-nontrivial.*

Proof. Immediate consequence of Theorem 196. ■

Finally, we state a result (Theorem 256), whose proof is an immediate consequence of the definition of P-consequence. However, before stating the theorem, let us introduce a definition:

DEFINITION 255 (Complementary Propositions). Let T be a C_{mp}-theory (in particular, a C-theory), and let α and β be formulas of the language of T. We say that α and β are *T-complementary* (or simply *complementary*) if there exists a formula γ of the language of T such that:

1. $T \vdash_{\mathsf{P}} \alpha$ and $T \vdash_{\mathsf{P}} \beta$
2. $T, \alpha \vdash_{\mathsf{P}} \gamma$ and $T, \beta \vdash_{\mathsf{P}} \neg\gamma$ (in particular, $\alpha \vdash_{\mathsf{P}} \gamma$ and $\beta \vdash_{\mathsf{P}} \neg\gamma$).

THEOREM 256. *If α and β are complementary theorems of a C_{mp}-theory T, and $\alpha \vdash_{\mathsf{P}} \gamma$ and $\beta \vdash_{\mathsf{P}} \neg\gamma$, then in general $\gamma \wedge \neg\gamma$ is not a theorem of T.*

Proof. Immediate, as a consequence of Theorem 196. ■

In other words, T is inconsistent from the point of view of classical logic, but it is P-nontrivial.

It should be emphasized, once again, that our way of characterizing complementarity does not mean that complementary propositions are always contradictory, for α and β above are not necessarily the negation of each other. However, since they are complementary propositions, we can derive from them (in classical logic) a contradiction. For example, 'x is a particle' is not the direct negation of 'x is a wave', but 'x is a particle' *entails* that 'x is not a wave'. This reading of complementarity as not indicating strict contradiction is in accordance with Bohr's view. As he notes (see [da Costa and Krause, 2004]):

> When we consider the well-known paradoxes that are encountered in the application of quantum theory to atomic structure, it is essential to remember, in this connection, that the properties of atoms are always obtained by observing their reactions under collisions or under the influence of radiation, and that the (...) limitation on the possibilities of measurement is directly related to the apparent contradictions that have been revealed in the discussion of the nature of light and of material particles. To emphasize that we *are not concerned here with real contradictions*, the author [Bohr himself] introduced in an earlier article the term 'complementarity'. [The italics is ours.]

Let us give a simple example of a situation involving a C_{mp}-theory. Suppose that our theory T is a fragment of quantum mechanics admitting Heisenberg's relations as a meaning postulate, and having as its underlying logic paraclassic logic. If α and β are two incompatible propositions according to Heisenberg's principle, we can interpret this principle as implying that α entails $\neg\beta$ (or that β entails $\neg\alpha$). So, even if we add α and β to T, we will be unable to derive, in T, $\alpha \wedge \beta$. Analogously, Pauli's exclusion principle has a similar interpretation to Heisenberg's.

As we have already noted, the basic feature of C_{mp}-theories is that, to make P-inferences, we suppose that some sets of statements we handle are consistent. In other words, C_{mp}-theories are closer to those theories scientists *actually* use in their everyday activity than those theories with the classical concept of deduction. That is, paraclassic logic — and paraconsistent logics in general — seem to fit more accurately the way scientists reason when they state and articulate their theories.

Generalization: the paralogic associated with a given logic

The technique used above to define the paraclassic logic associated with classical logic can be generalized to other logics $\mathcal{L}$. (This includes logics that have no negation symbol; but we will not discuss this case here.) The same point also applies to the concept of a C_{mp}-theory. More precisely, let $\mathcal{L}$ be a logic. It can be seen as a pair $\mathcal{L} = \langle \mathcal{F}, \vdash \rangle$, where $\mathcal{F}$ is a set of formulas of $\mathcal{L}$, and $\vdash \subseteq \mathcal{P}(\mathcal{F}) \times \mathcal{F}$ is the deduction relation of $\mathcal{L}$ (satisfying certain postulates depending on the particular logic $\mathcal{L}$). We define the $\mathsf{P}_{\mathcal{L}}$-logic associated with $\mathcal{L}$ (the 'paralogic' associated with $\mathcal{L}$) as follows:

Let $\mathcal{L}$ be a logic, which may be classical logic, intuitionistic logic, some paraconsistent logic or, in principle, any other logic. For simplicity, we suppose that the language of $\mathcal{L}$ has a symbol for negation, $\neg$. Then:

DEFINITION 257. A *theory* based on $\mathcal{L}$ (an *$\mathcal{L}$-theory*) is a set of formulas Γ of the language of $\mathcal{L}$ closed under $\vdash_{\mathcal{L}}$ (the symbol of deduction in $\mathcal{L}$). In other words, $\alpha \in \Gamma$, for every formula α such that $\Gamma \vdash_{\mathcal{L}} \alpha$.

DEFINITION 258. An $\mathcal{L}$-theory Γ is *$\mathcal{L}$-inconsistent* if there exists a formula α of the language of $\mathcal{L}$ such that $\Gamma \vdash_{\mathcal{L}} \alpha$ and $\Gamma \vdash_{\mathcal{L}} \neg\alpha$, where $\neg\alpha$ is the negation of α. Otherwise, Γ is *$\mathcal{L}$-consistent.*

DEFINITION 259. An $\mathcal{L}$-theory Γ is *$\mathcal{L}$-trivial* if $\Gamma \vdash_{\mathcal{L}} \alpha$, for every formula α of the language of $\mathcal{L}$. Otherwise, Γ is *$\mathcal{L}$-nontrivial.*

We can now define the $\mathsf{P}_{\mathcal{L}}$-logic associated with $\mathcal{L}$, whose language and syntactic concepts are those of $\mathcal{L}$. The only exception is the concept of deduction, which is introduced as follows: We say that α is a $\mathsf{P}_{\mathcal{L}}$-*syntactic consequence* of a set of formulas Γ, and write $\Gamma \vdash_{\mathsf{P}_{\mathcal{L}}} \alpha$, if and only if:

(1) $\alpha \in \Gamma$, or

(2) α is a provable formula of $\mathcal{L}$ (that is, $\vdash_{\mathcal{L}} \alpha$), or

(3) there exists $\Delta \subseteq \Gamma$ such that Δ is $\mathcal{L}$-nontrivial, and $\Delta \vdash_{\mathcal{L}} \alpha$.

As an example, we can consider the paraconsistent calculus C_1 as our logic $\mathcal{L}$. Then the paralogic associated with C_1 is a kind of 'para-paraconsistent' logic.

As a final remark, note that, sometimes, given a paraclassic theory T such that $T \vdash_{\mathsf{P}} \alpha$ and $T \vdash_{\mathsf{P}} \neg\alpha$, there exist *appropriate* propositions β and γ such that T can be replaced by a classical consistent theory T' in which $\beta \to \alpha$ and $\gamma \to \neg\alpha$ are theorems. If this happens, the logical difficulty can be, in principle, solved, and classical logic can be maintained.

7.4 Morality and Law

In its standard form, deontic logic can be taken as a kind of classical modal logic with the introduction of deontic operators: O for obligatory, P for permitted, V for prohibited, and F for indifferent. One of the interesting problems about deontic logic is that standard systems do not enable the existence of 'true' moral dilemmas. The latter are expressed by formulas of the form $O\alpha \wedge O\neg\alpha$, stating that both α and its negation are obligatory. This is due to the fact that, in standard deontic systems, $O\alpha \wedge O\neg\alpha \to O\beta$ is a theorem. Hence, a moral dilemma trivializes the whole system, in the sense that all formulas become obligatory. Other problems about standard deontic logics are created by Ross' paradox and the Good Samaritan paradox (see [Puga and da Costa, 1992]).

A standard solution to the problem about the nonexistence of moral dilemmas is simply to maintain that such dilemmas really cannot exist. However, for some philosophers, they are genuine possibilities (see [da Costa and Carnielli, 1986] for references). To accommodate this view, it is crucial to make room for moral dilemmas, but obviously without deontic triviality. Paraconsistent logics are excellent candidates for doing this.

The obligatory operator — as well as the other mentioned above – have various interpretations, the most common are in terms of *moral* obligations (and moral prohibitions etc.) and *legal* obligations (legal prohibitions etc.). If we write O_l and O_m for the legal and the moral obligations, respectively (the same notation applies to the other operators), it is important to study the principles $O_l\alpha \to P_m\alpha$ and $O_m\alpha \to P_l\alpha$, which seem to be quite 'natural'. To study these (and related) principles, some deontic paraconsistent logics have been introduced (see [Puga and da Costa, 1987]).

7.5 Philosophical significance of paraconsistent logic

After discussing the formal details of several paraconsistent systems, we can now examine some philosophical issues raised by the various constructions above as well as by paraconsistency more generally. Several issues need to be addressed. In particular, the issues of whether there is a true paraconsistent logic, and how to choose between such logics should be examined. In our view, the choice between such logics is ultimately a pragmatic and context-dependent matter, largely dependent on the details of the applications at hand. We will develop these points below.

The constructions we have just presented make clear that, at least at the level of mathematics, there are inconsistent but nontrivial theories. In other words, there are theories T in which both a formula α and its negation $\neg\alpha$ are theorems of T, and some formula β of T is *not* a theorem. In our view, this fact provides an important support for paraconsistency, since it shows that the attempt at accommodating inconsistencies by devising appropriate inconsistent but nontrivial theories is by no means empty or unrealizable. On the contrary, it provides a distinctive perspective on the issues under consideration. Instead of retaining classical logic, and avoiding the inconsistency by rejecting one or another of the premises which generate it — making more or less *ad hoc* moves — we retain the inconsistency, change the underlying logic to a paraconsistent one, and study the properties of the 'inconsistent object' so 'generated'. The important feature, as the paraconsistent set theory discussed above shows, is that these 'inconsistent objects' have certain determined properties and lack others: it is simply *not* the case that everything goes with regard to them. So, as opposed to what happens in the case of classical logic, there is a whole new domain of investigation determined by the formulation of paraconsistent logic: the domain of the inconsistent.

Now, the issue arises as to the status of the resulting inconsistent set theory: is it *true*? Well, can we say that there are true contradictions? The answer depends on several considerations. (1) What is the notion of truth used in this context? (2) What kinds of objects are we considering (only mathematical objects or do we include physical objects as well)? (3) What notion of existence is assumed? And how are ontological commitments to be spelled out? Of course, the examination of these issues involves particular philosophical moves, and we cannot do more here than consider them in fairly general terms. But we hope to say enough to make clear the approach we favor.

According to some authors, the answer to the question *Are there true contradictions?* is affirmative (see [Priest, 1987]). The examples given by Priest are the logical and semantic paradoxes, statements about moving objects (objects subject to change), and certain views in the foundations of mathematics. In order to articulate an 'ontic' or 'realist' view about true contradictions, Priest advocates (i) a strong notion of truth – truth *simpliciter* understood in the correspondence sense – (ii) a classical view about existence (as the range of bound variables), and (iii) an extended claim as to the domain of his theory — which incorporates both mathematical and physical objects. (Of course, in order to avoid triviality, Priest adopts a paraconsistent logic: see his system LP discussed in [Priest, 1987].) So, Priest's approach countenances classical views about truth and existence, and applies them to a wide-ranging domain. In our view, Priest's commitment to several doctrines is by no means fortuitous: in order to be adequately accommodated, inconsistencies require a whole package of logical and philosophical doctrines — indeed, a whole research program is involved. Of course, there are stronger and weaker programs; some are closer to classical proposals, some farther away.

It seems to us that, in retaining classical notions of truth and existence, Priest's proposal ends up committed to strong metaphysical views. Given the use of truth in the correspondence sense, and the claim that our assertions about the world (be it the 'empirical' or the 'mathematical' world) are to be true, Priest's view is *ipso facto* committed to all objects that are posited by these claims. In particular, his proposal seems to be committed to 'inconsistent objects' in the physical world: the objects to which our inconsistent but true physical theories refer. But how can their existence be established?

The argument to this effect assumes, of course, the classical account of ontological commitment: we are ontologically committed to whatever our bound variables range over. And in the case of inconsistent theories, this criterion leads us to postulate objects that both have and lack a given property (for instance, the liar sentence is both true and false, Russell set is both a member of itself and it isn't etc.). And the same goes for theories about the physical world.

In our view, this argument is not so conclusive. First, this criterion of ontological commitment is not independent of particular philosophical assumptions. It comes as part of a philosophical program — indeed, Quine's view — and it has built into it, as it were, a given logic: *classical* first-order logic. It goes without saying that, as such, it is at odds with Priest's own dialetheic approach, in which a paraconsistent logic is advocated. Moreover, Quine's criterion is *not* independent of logic: if we change the underlying logic of a given theory, we may change the entities we are quantifying over as well. This can be seen in several ways. For example, consider that we move to second-order logic. We are then allowed to quantify over properties and relations. As the result of second-order logic's strong expressive power, several mathematical theories can be better formulated in this setting (in particular, as is well known, arithmetic and analysis are categorical in a second-order context). Because of this, several nominalist proposals (such as those developed in [Field, 1980] and [Hellman, 1989]) have adopted second-order logic as part of their nominalization strategies of science and mathematics. The idea is that, by increasing the strength of the logic, we can decrease our ontological commitments. Secondly, using paraconsistent logic, we are allowed to quantify over certain constructions (such as Russell set) that are impossible in classical logic, given the latter's identification of inconsistency and triviality.

Our point here is that Quine's slogan — to be is to be the value of a variable — can only have any force once a particular logic is admitted. Quine knows that, of course. The problem is that his view assumes a logic (classical fist-order logic) that is not the most adequate to deal with inconsistencies in a heuristically fruitful way.

We suggest a different way of addressing the inconsistency issue. We may well explore the rich representational devices allowed by the use of paraconsistency in inconsistent domains, but withholding any claim to the effect that there are 'inconsistent objects' in reality. Whether the world is indeed 'inconsistent' — assuming there is a sensible formulation of this claim — is something we would rather be *agnostic* about. Just as empiricists (such as van Fraassen [van Fraassen, 1980]) are agnostic about the existence of unobservable entities in science, we are agnostic about the existence of true contradictions in nature. And one of the reasons in support of this claim is an *underdetermination* argument. Given the hierarchy of paraconsistent C-logics discussed above, there are always infinitely many paraconsistent logics that can be used to accommodate a given ‘phenomenon’ — whether it is an ‘inconsistent’ reasoning or an ‘inconsistent’ theory (see [da Costa *et al.*, 1998c]). Which of these paraconsistent logics reflects *the* logic of the world? There is no argument based on purely observational terms that could settle the issue. We can select, of course, one of these logics on *pragmatic* grounds, but these grounds are certainly not enough to establish a substantive claim about the world. For instance, if one of these logics makes the modeling of the inconsistency in question easier, why should this be taken as a reason for this logic to be *true*? Simplicity may well be a sensible criterion to adopt on pragmatic grounds, but the claim that a logic selected on this basis is (likely to be) true, is to confuse pragmatic and epistemic considerations. Why should the world conform to our cognitive limitations? Of course, it might well do. But to establish this claim demands an argument that goes beyond what we observe: it requires a metaphysical commitment to the simplicity of reality. And, to a certain extent, this is as ‘strong’ as the claim that there are true contradictions. After all, both make substantial assertions about the world that transcend empirical observation. Both are metaphysical claims.

In our view, an alternative program of interpretation of inconsistencies can be devised in which no commitment to this kind of metaphysics is required. The idea is first to avoid the claim that inconsistent theories are *true*; they are *partially true* (or *quasi-true*) at best. As discussed above, the notion of partial truth receives a formal treatment. But for our needs, it suffices to recall that a sentence α is partially true if it models adequately only part of a given domain D, leaving open the ‘complete’ description of the latter. (Of course, in a precise sense, α is consistent with a true description of D.) In claiming that with regard to inconsistent theories all we need is to determine their partial truth, we are in position to provide a ‘formal underpinning’ to our agnosticism with regard to true contradictions. We are suggesting a change in the notion of truth to the weaker notion of partial truth, withholding then the commitment to ‘inconsistent objects’. After all, several ($\mathfrak{A}$-normal) structures describe the domain under consideration, and inconsistent objects are not countenanced in all of them.[30] Moreover, we also suggest a revision of Quine’s slogan about ontological commitment, making explicit its dependence on the underlying logic. In this way, it becomes clear that this slogan is *not* the only criterion to adjudicate between alternative logics.

However, if we end up not being committed to an ontology of actual inconsistent physical objects, are we committed to inconsistent *mathematical* entities? This depends, of course, on how we interpret the relevant mathematical theory. Does it provide a *true* description of the inconsistent mathematical ‘world’? Again, if all we claim is to have described this ‘world’ in at best partially true terms, we will not be committed to those entities. There is, of course, a whole story to be told here, but for the present purposes, it suffices to note that it is possible to provide an entirely syntactic formulation of paraconsistent set theory, in which various inconsistent theories can be embedded, such that the only commitment is to a countable language. Thus, in a certain sense, we do not require a special commitment to mathematical objects either.

---

[30] Descartes once remarked (in his *Principles*, iv, 204) that, ‘with regard to those things that our senses cannot perceive, it suffices to explicate how they can be’. We would say the same with regard to inconsistent theories. The notion of partial truth allows us to accommodate formally this point, since partial truth is strictly weaker than truth, and does not commit us to anything *beyond* the assertion that certain structures are *possible*, given some paraconsistent logic.

So, the 'package' we suggest to accommodate inconsistencies is characterized by (1) the claim that inconsistent theories are *partially true* at best; (2) an agnosticism with regard to the existence of true contradictions and a nominalism about inconsistent mathematical entities; and (3) a re-evaluation of Quine's view about ontological commitment, emphasizing its dependence on the underlying logic.

In our view, the striking feature of this 'package' is its logical pluralism, on the one hand, and the fact that it allows us to make sense of the several uses and applications of paraconsistent logic with no commitment to actual 'inconsistent objects'. The logical pluralism derives from claim (3) above. Depending on the domain we are considering, different kinds of logic may be appropriate. For instance, if we want to model constructive features in mathematical reasoning, an intuitionistic logic is the best alternative; if we are concerned with inconsistent bits of information, the use of a paraconsistent logic is the natural option. In particular, we do not reject classical logic: it has its own domains and applications. To this extent, while dealing with distinct domains, paraconsistent logic and classical logic are complementary rather than rivals. (They become rivals only when applied to the *same* domain. The rivalry derives from the fact that they provide different accounts of the logical connectives.)

But in applying paraconsistent logic, as we have done when formulating the theory of the Russell set and other inconsistent objects, we do not have to be committed to the existence of 'inconsistent entities' — this is the point of our claims (1) and (2) above. We can always use the resources provided by this logic only to help us draw consequences from inconsistent theories without triviality, but with no commitment to the *truth* of the theory in question; it can be at best partially true.

In this way, a noncommittal (agnostic) interpretation of paraconsistency can be presented, making sense of the application of paraconsistent logic articulated here. 'Inconsistent objects', either mathematical or physical, can be accommodated without requiring an ontology which includes them. In particular, inconsistent mathematical theories can be studied in the context of paraconsistent logic, but it is not necessary to countenance the existence of the entities the theories are taken to be about.

8 CONCLUDING REMARKS

We noted above that it is nearly impossible for anyone to follow all the literature on paraconsistent logics. We have left untouched numerous important topics here. Nowadays, 'paraconsistency' has become a field of knowledge so wide that only events such as World Congresses on Paraconsistency can begin to do justice to it.

As a curiosity, a search for 'paraconsistent logic' on Google (`www.google.com`) on October 13, 2005 yielded 61,600 entries! We should stop now.

ACKNOWLEDGMENTS

We would like to thank W. Carnielli, J. M. Abe, M. Barreto, A. Costa Leite and J. -Y. Béziau, M. Guillaume, Q. Zhang, J. Marcos and I. D'Ottaviano for sending us updated information on their works on paraconsistency. We have not been able to incorporate all of their helpful comments and suggestions in this work. But we hope to do that in our forthcoming book.

BIBLIOGRAPHY

[Abar, 1985] C. A. A. P. Abar. *Descrição e Paraconsistência*, Doctor Thesis, Ponthifical Catholic University of São Paulo, S. Paulo, 1985.

[Abe, 1992] J. M. Abe. *Fundamentos da Lógica Anotada*, Thesis, University of São Paulo, 1992.

[Abe and da Silva Filho, 1998] J. M. Abe and J. I. da Silva Filho. Inconsistency and electronic circuits, *Proceedings of The International ICSC Symposium on Engineering of Intelligent Systems* (EIS'98), Volume 3, Artificial Intelligence, Editor: E. Alpaydin, ICSC Academic Press International Computer Science Conventions Canada/Switzerland, 1998, pp. 191-197.

[Abe and da Silva Filho, 2003] J. M. Abe and J. I. da Silva Filho. 'Manipulating Conflicts and Uncertainties in Robotics', *Multiple-Valued Logic and Soft Computing*, V.9, pp. 147-169, 2003.

[Akama, 2002] S. Akama. Nelson's paraconsistent logics, *Logic and Logical Philosophy*, 1999 (published in 2002).

[Akama and Abe, 2001] S. Akama and J. M. Abe. The role of inconsistency in information systems, *Proc. of the 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI'2001)*, pp. 355-360, Orlando, USA, 2001.

[Ávila, 1996] B. C. Ávila. *Uma Abordagem Paraconsistente Baseada em Lógica Evidencial para Tratar Exceções em Sistemas de Frames com Múltipla Herança*, Thesis, University of São Paulo, 1996.

[Ávila and Abe, 1997] B. C. Ávila, J. M. Abe and J. P. A. Prado. Reasoning in paraconsistent frame systems, *The Second International Workshop on CSCW in Design*, P. Siriruchatapong Z. Lin & J. P. Barthes (Eds), International Academic Publishers, Bangkok, Thailand, pp. 239-244, 1997.

[Alves, 1976] E. H. Alves. *Lógica e Inconsistência: um estudo dos cálculos* C_n $1 \leq n \leq \omega$, Master Dissertation, Universidade Estadual de Campinas, 1976.

[Alves, 1984] E. H. Alves. Paraconsistent logic and model theory, *Studia Logica* **43**(1/2), 1984, pp. 17-32.

[Alves and Moura, 1978] E. H. Alves and J. E. Moura. On some higher order paraconsistent calculi, in A. I. Arruda, N. C. A. da Costa and R. Chuaqui (eds.), *Mathematical Logic: Proceedings of the First Brazilian Conferece*, Marcel Dekker, 1978, pp. 1-8.

[Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap, Jr. *Entailment, the Logic of Relavance and Necessity*, Princeton Un. Press, 1976.

[Arruda, 1964] A. I. Arruda. *Considerações sobre os sistemas formais* $\mathcal{NF}_n$, Thesis, Universidade Federal do Paraná, 1964.

[Arruda, 1977] A. I. Arruda. On the imaginary logic of N. A. Vasil'év, in Arruda, A. I., da Costa, N. C. A. and Chuaqui, R. (eds.), *Non-Classical Logics, Model Theory and Computability*, North Holland, 1977, pp. 3-22.

[Arruda, 1980] A. I. Arruda. A survey of paraconsistent logic, in Arruda, A. I., Chuaqui, R. and da Costa, N. C. A. (eds.), *Mathematical Logic in Latin America*, North-Holland, 1980, pp. 3-41.

[Arruda, 1989] A. I. Arruda. Aspects of the historical development of paraconsistent logic, in Priest, G., Routley, R. and Norman, J., (eds.), *Paraconsistent Logic*, Philosophia Verlag, 1989, pp. 99-130.

[Arruda, 1990] A¿ I. Arruda. *N. A. Vasiliev e a Lógica Paraconsistente*, Centro de Lógica, Epistemologia e História da Ciência, State Univesity of Campinas, Col. CLE v. 7, 1990.

[Arruda and da Costa, 1966] A. I. Arruda and N. C. A. da Costa. O paradoxo de Curry-Moh Shaw-Kwei, *Boletim da Sociedade de Matemática de São Paulo* **18**, 1966, pp. 83-89.

[Arruda and da Costa, 1970] A. I. Arruda and N. C. A. da Costa. Sur le schéma de la séparation, *Nagoya Math. Journal* **38**, 1970, pp. 71-84.

[Arruda and da Costa, 1977] A. I. Arruda and N. C. A. da Costa. Une sémantique pour le calcul C_1. *C. R. Acad. Sc. Paris* **284** (A), 1977, 279-282.

[Arruda and da Costa, 1984] A. I. Arruda and N. C. A. da Costa. On the relevant systems P and $P^\star$ and some related systems, *Studia Logica* **43**, 1984, pp. 33-49.

[Asenjo, 1966] F. G. Asenjo. A calculus of antinomies, *Notre Dame J. of Formal Logic* **XVI**, 1966., pp. 103-105.

[Barreto, 2002] M. M. G. Barreto. Handling Paraconsistent Information with Possibilistic Logic, *Coleção Documentos*, Série Lógica e Teoria da Ciência, Instituto de Estudos Avançados, Universidade de São Paulo, 2002.

[Barreto, 2004] M. M. G. Barreto. Handling Paraconsistent Knowledge with Possibilistic Resolution. To appear in *Coleção Documentos*, Série Lógica e Teoria da Ciência, Instituto de Estudos Avançados, Universidade de São Paulo, 2004.

[Barreto and Ebecken, 1998] M. M. G. Barreto and N. F. F. Ebecken. An Intelligent System for Helping Petroleum Industry Risk Management Decision Making, *Proceedings of the Rio Oil & Gas Expo and Conference*, Rio de Janeiro, RJ, 1998.

[Barreto and Ebecken, 2001] M. M. G. Barreto and N. F. F. Ebecken. Using Paraconsistent and Fuzzy Logics to Build up Inconsistent KBS, *Proceedings of the 5th World Multi Conference on Systems, Cybernetics and Informatics* (SCI2001) and the *7th International Conference on Information Systems, Analysis and Synthesis* (ISAS2001), Orlando, USA, 2001.

[Barreto and Ebecken, 2002] M. M. G. Barreto, N. F. F. Ebecken and N. C. A. da Costa. The Defuzzification Process of a Paraconsistent Model, *Coleo Documentos*, Série Lógica e Teoria da Ciência, Instituto de Estudos Avançados, Universidade de São Paulo, Número 42, 2002.

[Batens, 1980] D. Batens. Paraconsistent extensional propositional logics, *Logique et Analyse* **90-91**, pp. 195-234, 1980.

[Batens, 1999] D. Batens. Inconsistency-adaptive logics, in e. Orłowska (ed.), *Logic at Work. Essays dedicated to the memory of Elena Rasiowa*, Springer-Verlag, pp. 445-472, 1999.

[Batens, 2000] D. Batens. Minimally abnormal models in some adaptive logics, *Synthese* **125**, pp. 5-18, 2000.

[Béziau, 1990] J.-Y. Béziau. Logiques construites suivant les méthodes de da Costa, *Logique et Analyse* **131-132**, pp. 259-272, 1990.

[Béziau, 1995a] J.-Y. Béziau. Negation: what it is and what it is not, *Boletim da Sociedade Paranaense de Matemática* **15** (1/2), pp. 37-43, 1995.

[Béziau, 1995b] J.-Y. Béziau. Universal logic, in Childers, T. and Majer, O. (eds.), *Proceedings of the 8th International Colloquium Logica'94*, Praga, Philosophia, pp. 73-93, 1995.

[Béziau, 1997] J.-Y. Béziau. La logique paraconsistante, Appendix A of da Costa, N. C. A., *Logiques Classiques et Non Classiques*, Paris, Masson, pp. 237-255, 1997.

[Béziau, 2002a] J.-Y. Béziau. Are paraconsistent negations negations? In *Paraconsistency: the logical way to the inconsistent*, W. Carnielli et al. (eds), Marcel Dekker, New-York, pp. 465-486, 2002.

[Béziau, 2002b] J.-Y. Béziau. S5 is a paraconsistent logic and so is first-order classical logic, *Logical Investigations* **9**, pp. 301-309, 2002.

[Béziau, 2003] J.-Y. Béziau. Talk given at Federal University of Santa Catarina, 2003.

[Blair and Subrahmanian, 1989a] H. A. Blair and V. S. Subrahmanian. Paraconsistent Logic Programming, *Theoretical Computer Science* **68**, pp. 135-154, 1989.

[Blair and Subrahmanian, 1988b] H. A. Blair, H. A. and Subrahmanian, V. S.,'Paraconsistent foundations of logic programming', *J. Non Classical Logic* **5**, pp. 44-73, 1989.

[Blair and Subrahmanian, 1989c] H. A. Blair and V. S. Subrahmanian. Paraconsistent logic programming, in Proceedings of the 7th Conference on Foundations of Software Technology and Theoretical Computer Science **287**, 1987, pp. 340-360. Extended version in *Theoretical Computer Science* **68**, 1989, pp. 135-154.

[Bell, 1988] J. L. Bell. Infinitesimals', *Synthese* **75**, pp. 285-315, 1988.

[Bell, 1995] J. L. Bell. Infinitesimals and the continuum', *Mathematical Intelligencer* **17**, pp. 55-57, 1995.

[Beller, 1992] M. Beller. The birth of Bohr's complementarity: the context and the dialogues', *Stud. Hist. Phil. Sci.* **23** (1), 147-180, 1992.

[Bibel, 1982] W. Bibel. *Automated Theorem Proving*, Friedr. Vieweg & Sohn, Braunschweig, Wiesbaden, 1982.

[Bibel, 1993] W. Bibel. *Deduction: automated logic*, Academic Press, 1993.

[Bobenrieth, 1996] A. Bobenrieth Miserda. *Inconsistencias, ?por que no?*, Bogotá, Cocultura, 1996.

[Borel, 1950] E. Borel. *Leçons sur la théorie des functions*, Gauthier-Villars, 1950.

[Bourbaki, 1968] N. Bourbaki. *Theory of Sets*, Paris and Reading, MA, Hermann and Addison-Wesley, 1968.

[Buchsbaum and Pequeno, 1993] A. Buchsbaum and T. Pequeno. A reasoning method for a paraconsistent logic', *Studia Logica* **52** (2), pp. 281-289, 1993.

[Buchsbaum *et al.*, 2001] A. Buschbaum, T. Pequeno and M. Pequeno. A Positive Formalization for the Notion of Pragmatic Truth, *Proceedings of 2001 International Conference on Artificial Intelligence* - 25 a 28 de junho de 2001 - Las Vegas, Nevada, USA, 2001.

[Buchsbaum *et al.*, 2004] A. Buschbaum, T. Pequeno and M. Pequeno. The Logical Expression of Reasoning'. In: *New Trends in the Foundations of Science* (Papers dedicated to the 80th birthday of Patrick Suppes presented in Florianópolis, Brazil, April 22-23, 2003), Collection Logic, Epistemology and the Unity of Science, J. -Y. Béziau and D. Krause (eds.), to appear.

[Bueno, 1997] O. Bueno. Empirical Adequacy: A Partial Structures Approach', *Studies in History and Philosophy of Science* **28**, pp. 585-610, 1997.

[Bunder, 1980] M. Bunder. A new hierarchy of paraconsistent logics, in A. I. Arruda, N. C. A. da Costa and A. M. Sette (eds.), *Proceedings of the Third Brazilian Conference on Mathematical Logic*, Sociedade Brasileira de Lógica, São Paulo, pp. 13-22, 1980.

[Bunder, 1983] M. Bunder. On Arruda and da Costa's logics J_1 to J_5, *J. Non-Classical Logic* **2**, pp. 43-48, 1983.

[Caiero and de Souza, 1997] R. da C. Caiero and E. G. de Souza. A new paraconsistent set theory: ML_1, *Logique et Analyse* **157**, pp. 115-141, 1997.

[Carnielli, 2002] W. A. Carnielli, M. E. Coniglio and I. M. L. D'Ottaviano. *Paraconsistency: the logical way to the inconcistent*, Proceedings of the Second World Congress on Paraconsistency, held in São Paulo, Marcel Dekker, 2002.

[Carnielli, 1987] W. A. Carnielli. Systematization of finite many-valued logics through the method of tableaux', *J. Symb. Logic* **52** (2), pp. 473-493, 1987.

[Carnielli and Alcântara, 1984] W. A. Carnielli and L. P. Alcântara. Paraconsistent algebras, *Studia Logica* **43**, pp. 79-88, 1984.

[Carnielli and Marcos, 2002] W. A. Carnielli and J. Marcos. A Taxonomy of C-Systems, in Carnielli, W. A., Coniglio, M. E. and D'Ottaviano, I. M. L. eds., *Paraconsistency - the Logical Way to the Inconsistent.* Lecture Notes in Pure and Applied Mathematics, Vol. 228, pp. 1-94. Marcel Dekker, New York, 2002.

[Carnielli *et al.*, 2006] W. A. Carnielli, M. E. Coniglio and J. Marcos. Logics of Formal Inconsistency'. To appear in the Handbook of Philosophical Logic, Kluwer Academic Publishers (eds. D. Gabbay and F. Guenthner).

[Carnielli *et al.*, to appear] W. A. Carnielli, M. E. Coniglio and A. Costa-Leite. Paraconsistency and knowability', to appear.

[Correa *et al.*, 1993] M. Correa, A. Buchsbaum and T. Pequeno. Sensible inconsistent reasoning, *Technical Notes of The AAAI Workshop on Automated Deduction for Non Standart Logics*, North Carolina, 1993.

[Church, 1956] A. Church. *Introduction to Mathematical Logic*, Vol. I, Princeton Un. Press, 1956.

[Church, 1974] A. Church. Set theory with a universal set, in Henkin, L. (ed.), *Proceedings of the Tarski Symposium*, Providence, American Mathematical Society, pp. 297-308, 1974.

[da Costa, 1958] N. C. A. da Costa. Nota sobre o conceito de contradio', *Anuário da Sociedade Paranaense de Matemática* **1**, pp. 6-8, 1958.

[da Costa, 1959] N. C. A. da Costa. Observações sobre o conceito de existência em matemática', *Anuário da Sociedade Paranaense de Matemática* **2**, pp. 16-19, 1959.

[da Costa, 1963a] N. C. A. da Costa. *Sistemas Formais Inconsistentes*, Rio de Janeiro, NEPE, 1963. Reprinted by Editora da UFPR, Curitiba, 1993.

[da Costa, 1963b] N. C. A. da Costa. Calculs propositionels pour les systèmes formales inconsistants, *C. R. Acad. Sc. Paris* **257**, 1963, pp. 3790-3793, 1964.

[da Costa, 1964a] N. C. A. da Costa. Calculs de prédicats pour les systèmes formales inconsistants, *C. R. Acad. Sc. Paris* **258**, 1964, pp. 27-29, 1964.

[da Costa, 1964b] N. C. A. da Costa. Sur un système inconsistant de la théorie des ensembles', *C. R. Acad. Sc. Paris* **258**, pp. 3144-3147, 1964.

[da Costa, 1966a] N. C. A. da Costa. *Álgebras de Curry*, São Paulo, 1966.

[da Costa, 1966b] N. C. A. da Costa. Opérations non-monotones dans les treillis, *C. R. Acad. Sc. Paris* **263A**, pp429-432, 1966.

[da Costa, 1969] N. C. A. da Costa. 'On a set theory suggested by Ehresmann and Dedecker', *Proceedings of the Japan Academy of Sciences* **45**, pp. 880-888, 1969.

[da Costa, 1974] N. C. A. da Costa. On the theory of inconsistent formal systems, *Notre Dame Journal of Formal Logic* **XV** (4), pp. 497-510, 1974.

[da Costa, 1975] N. C. A. da Costa. Remarks on Jaśkowski's discussive logic, *Rep. Math. Logic* **4**, pp. 7-16, 1975.

[da Costa, 1980] N. C. A. da Costa. *Ensaio sobre os Fundamentos da Lógica*, Hucitec-EdUSP, S. Paulo, 1980.

[da Costa, 1982] N. C. A. da Costa. The philosophical import of paraconsistent logic, *The Journal of Non-Classical Logic*, **1**, pp. 1-19, 1982.

[da Costa, 1986a] N. C. A. da Costa. On paraconsistent set theory, *Logique et Analyse* **115**, pp. 361-371, 1986.

[da Costa, 1986b] N. C. A. da Costa. Pragmatic Probability', *Erkenntnis* **25**, pp. 141-162, 1986.

[da Costa, 1989] N. C. A. da Costa. Mathematics and Paraconsistency (in Portuguese), *Monografias da Sociedade Paranaense de Matemática* **7**, 1989.

[da Costa, 1993] N. C. A. da Costa. *Lógica Indutiva e Probabilidade*, Hucitec-EdUSP, 2nd. ed., 1993.

[da Costa, 1997] N. C. A. da Costa. *Logique Classique et Non-Classique*, Paris, Masson, 1997 (French translation of [da Costa, 1980], with Preface and Appendices by J. -Y. Béziau).

[da Costa and Alcântara, 1982] N. C. A. da Costa, L. P. and Alcântara. On paraconsistent set theories, *Relatório Interno* 215, IMECC-Unicamp, 1982.

[da Costa and Alcântara, 1986] N. C. A. da Costa, L. P. and Alcântara. A note on type theory, *Comptes rendus de l'Académie bulgare des Sciences* **39** (10), pp. 5-7, 1986.

[da Costa and Alcântara, 1991] N. C. A. da Costa and L. P. Alcântara. On paraconsistent set theories, *Bol. Soc. Paran. Mat.* **12/13** (1/2), pp. 78-81, 1991/1992.

[da Costa and Alves, 1977] N. C. A. da Costa and E. H. Alves. A semantical analysis of the calculi C_n, *Notre Dame J. of Formal Logic* **18** (4), pp. 621-630, 1977.

[da Costa and Carnielli, 1986] N. C. A. da Costa and W. A. Carnielli. On paraconsistent deontic logic, *Philosophia* **16** (3/4), pp. 293-305, 1986.

[da Costa and Chuaqui, 1988] N. C. A. da Costa and R. Chuaqui. On Suppes' set theoretical predicates, *Erkenntnis* **29**, pp. 95-112, 1988.

[da Costa and Doria, 1995] N. C. A. da Costa and F. A. Doria. On Jaśkowski's discussive logic, *Studia Logica* **54**, pp. 33-60, 1995.

[da Costa and Dubikajtis, 1968] N. C. A. da Costa and L. Dubikajtis. Sur la logique discoursive de Jaśkowski, *Bull. Acad. Polonaise des Sciences* **26**, pp. 551-557, 1968.

[da Costa and Dubikajtis, 1977] N. C. A. da Costa and L. Dubikajtis. On Jaśkowski's discussive logic, in Arruda, A. Y., da Costa, N. C. A. and Chuaqui, R. (eds.), *Non-Classical Logic, Model Theory and Computability*, North-Holland, pp. 37-56, 1977.

[da Costa and French, 1989a] N. C. A. da Costa and S. French. Pragmatic truth and the logic of induction, *The British Journal for the Philosophy of Science* **40**, pp. 333-356, 1989.

[da Costa and French, 1989b] N. C. A. da Costa and S. French. In contradiction. Review of G. Priest, 'In Contradiction: A Study of the Transconsistent', *Phil. Quarterly* **39**, pp. 498-502. 1989.

[da Costa and French, 1990] N. C. A. da Costa and S. French. The model-theoretic approach in the philosophy of science, *Philosophy of Science* **57**, pp. 248-265, 1990.

[da Costa and French, 2003a] N. C. A. da Costa and S. French. Towards an acceptable theory of acceptance: partial structures, inconsistency and dorrespondence, in S. French and H. Kamminga (eds.), *Correspondence, Invariance and Heuristics*, Dordrecht, Kluwer Academic Publishers, pp. 137-158, 1993.

[da Costa and French, 1993b] N. C. A. da Costa and S. French. A model theoretic approach to 'natural reasoning', *International Studies in the Philosophy of Science* **7**, pp. 177-190, 1993.

[da Costa and French, 2003] N. C. A. da Costa and S. French. *Partial Truth: a unitary approach to models and scientific reasoning*, Oxford University Press, 2003.

[da Costa and Krause, 2002] N. C. A. da Costa and D. Krause. An inductive annotated logic, in W. Carnielli, M. Cogniglio and I. M. L. D'Ottaviano (eds.), *Paraconsistency: the Logical Way to the Inconsistent*, Proceedings do II World Congress on Paraconsistency, Juquehy, SP, May 2000 (Marcel Dekker, New York), pp. 213-225, 2002.

[da Costa and Krause, 2004] N. C. A. da Costa and D. Krause. Complementarity and paraconsistency, in S. Rahman; J. Symons; D. M. Gabbay; J. -P. van Bendegen (eds.), *Logic, Epistemology, and the Unity of Science*, forthcoming by Kluwer Ac. Press, 2004.

[da Costa and Krause, 2004a] N. C. A. da Costa and D. Krause. The logic of complementarity, in G. Heinzmann (ed.), *The Age of Alternative Logics: assessing philosophy of logic and mathematics today*, forthcoming. (http://philsci-archive.pitt.edu/archive/00001559)

[da Costa and Krause, 2004b] N. C. A. da Costa and D. Krause. Remarks on the applications of paraconsistent logic to physics, forthcoming. (http://philsci-archive.pitt.edu/archive/00001566)

[da Costa and Marconi, 1989] N. C. A. da Costa and D. Marconi. An overview of paraconsistent logic in the 80s, *J. Non-Classical Logic* **6** (1), pp. 5-32, 1989.

[da Costa and Sette, 1969] N. C. A. da Costa and A. M. Sette. Les algèbres C_ω, *C. R. Acad. Sci. Paris* **268**, pp. 1011-1014, 1969.

[da Costa and Subrahmanian, 1989] N. C. A. da Costa and V. S. Subrahmanian. Paraconsistent logics as a formalism for reasoning about inconsistent knowledge bases, *Artificial Intelligence in Medicine* **1**, pp. 167-174, 1989.

[da Costa and Vernengo, 1999] N. C. A. da Costa R. J. y Vernengo. Sobre algunas lógicas paraclássicas y el análisis del razonamiento jurídico, *Doxa* **19**, pp. 183-200, 1999.

[da Costa and Wolf, 1980] N. C. A. da Costa and R. G. Wolf. Studies in paraconsistent logic I: The dialectical principle of the unity of opposites, *Philosophia (Philosophical Quarterly of Israel)* **9**, no. 2, pp. 189-217, 1980.

[da Costa and Wolf, 1985] N. C. A. da Costa and R. G. Wolf. Studies in paraconsistent logic II: quantification and the unity of opposites, *Rev. Comobiana de Matemáticas* **19**, pp. 56-67, 1985.

[da Costa *et al.*, 1991] N. C. A. da Costa, J. M. Abe and V. S. Subrahamanian. Remarks on annotated logic, *Zeitschr. f. math. Logik und Grundlagen d. Math.* **37**, pp. 561-570, 1991.

[da Costa *et al.*, 1995] N. C. A. da Costa, J.-Y. Béziau and O. Bueno. Paraconsistent logic in a historical perspective, *Logique et Analyse* **150-151-152**, pp. 111-125, 1995.

[da Costa *et al.*, 1998a] N. C. A. da Costa, J.-Y. Béziau and O. Bueno, O., *Elements of Paraconsistent Set Theory* (in Portuguese), Campinas, Centro de Lógica, Epistemologia e História da Ciência (Coleo CLE 23), 1998.

[da Costa *et al.*, 2004] N. C. A. da Costa, O. Bueno and A. G. Volkov. Outlines on a paraconsistent category theory, in P. Weingartner (ed.), *Alternative Logics: do sciences need them?* Springer, pp. 95-114, 2004.

[da Costa *et al.*, 1998b] N. C. A. da Costa, O. Bueno and S. French. The logic of pragmatic truth, *Journal of Philosophical Logic* **27**, pp6́03-620, 1998.

[da Costa *et al.*, 1998c] N. C. A. da Costa, O. Bueno and S. French. Is there a Zande logic? *History and Philosophy of Logic* **19**, pp. 41-54, 1998.

[da Costa *et al.*, 1996] N. C. A. da Costa, R. Chuaqui and O. Bueno. The logic of quasi-truth, *Notas se la Sociedad de Matematica de Chile* **XV**, pp. 7-26, 1996.

[da Costa *et al.*, 1990] N. C. A. da Costa, L. J. Henschen, J. J. Lu, and V. S. Subrahamanian. Automatic theorem proving in paraconsistent logics: foundations and implementation, in Proceedings of the 10th International Conference on Automated Deduction, *Springer Lecture Notes in Computer Science* **449**, pp. 72-86, 1990.

[da Costa *et al.*, 1991] N. C. A. da Costa, V. S. Subrahamanian and C. Vago. The paraconsisten logic $\mathsf{P}\mathcal{T}$, *Zeitschr. f. math. Logik und Grundlagen d. Math.* **37**, pp. 139-148, 1991.

[Dalla Chiara and di Francia, 1981] M. L. Dalla Chiara amd G. Toraldo di Francia. *Le Teorie Fisiche: un'analisi formale*, Torino, Boringhieri, 1981.

[Dalla Chiara and Giuntini, 1989] M. L. Dalla Chiara and R. Giuntini. Paraconsistent quantum logic, *Foundations of Physics* **19**, 1989, pp. 891-904.

[Dalla Chiara and Giuntini, 2000] M. L. Dalla Chiara and R. Giuntini. Paraconsistent ideas in quantum logic, *Synthese* **125**, pp. 55-68, 2000.

[Dalla Chiara and Giuntini, 2001] M. L. Dalla Chiara and R. Giuntini. Quantum logic, in `http://xxx.lanl.gov/list/quant-ph/0101`.

[D'Ottaviano, 1982] I. M. L. D'Ottaviano. *Sobre uma Teoria de Modelos Trivalente*, Ph.D. Dissertation, State University of Campinas, 1982.

[D'Ottaviano, 1985a] I. M. L. D'Ottaviano. The completeness and compactness of a three valued first-order logic', *Revista Colombiana de Matemácicas* **15**, pp. 31-42, 1995. s

[D'Ottaviano, 1985b] I. M. L. D'Ottaviano. The model extension theorems for J_3-theories, in C. A. di Prisco (ed.), *Methods in Mathematical Logic*, Springer, pp. 157-183, 1985.

[D'Ottaviano, 1990] I. M. L. D'Ottaviano. On the development of paraconsistent logic and da Costa's work, *J. Non-Classical Logic* **7** (1/2), pp. 9-72, 1990.

[D'Ottaviano and da Costa, 1970] I. M. L. D'Ottaviano and N. C. A. da Costa. Sur un problème de Jaśkowski, *Comptes Rendus de l'Académie des Sciences de Paris* **270A**, pp. 1349-1353, 1970.

[Dubois and Prade, 1988] D. Dubois and H. Prade. *Possibility theory: an approach to computerized processing of uncertainty*, Plenum Press, 1988.

[Ellentuck, 1975] E. Ellentuck. Gödel's square axioms for the continuum', *Math. Ann.* **216**, pp. 29-33, 1975.

[Fidel, 1977] M. Fidel. The decidability of the calculi C_n, *Reports on Mathematical Logic* **8**, pp. 31-40, 1977.

[Field, 1980] H. Field. *Science without Numbers*, Princeton, N.J., Princeton University Press, 1980.

[French and Krause, to appear] S. French and D. Krause. *Identity and Individuality in Modern Physics*, to appear.

[Furmanowski, 1975] T. Furmanowski. Remarks on discussive propositional calculus, *Studia Logica* **34**, pp. 39-43, 1975.

[Gallin, 1975] D. Gallin. *Intensional and Higher-Order Logic*, Amsterdam, North-Holland, 1975.

[Gonseth, 1974] F. Gonseth. *Les Mathématiques et la Réalité: essay sur la méthode axiomatique*, Albert Blanchard, 1936 (1974).

[Grana, 1983] N. Grana. *Logica Paraconsistente*, Loffredo Editore, Napoli, 1983.

[Grana, 1990a] N. Grana. *Logica Deontica Paraconsistente*, Liguori Editore, Napoli, 1990.

[Grana, 1990b] N. Grana. *Sulla Teoria delle Valutazioni di da Costa*, Liguori Editore, Napoli, 1990.

[Grana, 1990c] N. Grana. *Contradizione e Incompletezza*, Liguori Editore, Napoli, 1990.

[Grana, 2001] N. Grana. *Epistemologia della Matematica: ontologia, verità, valutazioni*, L'Orientale Editrice, Napoli, 2001.

[Granger, 1998] G.-G. Granger. *L'irrationnel*, Odile Jacob, 1998.

[Grant and Subrahmanian, 2000] J. Grant and V. S. Subrahmanian. 'Applications of paraconsistency in data and knowlwedge basis', *Synthese* **125**, pp. 121-132, 2000.

[Guillaume, 1960] M. Guillaume. Recherches sur le symbole de Hilbert, Clermond-Ferrand, 1960.

[Guillaume, 1996] M. Guillaume. Regard en arrière sur quinze annés de coopération douce avec l'écolebrézilienne de logique paraconsistante' *Logique et Analyse* **153-154**, pp. 5-14, 1996.

[Haack, 1974] S. Haack. *Deviant Logic*, Cambridge Un. Press, 1974.

[Hatcher, 1982] W. S. Hatcher. *The Logical Foundations of Mathematics*, Toronto, Pergamon Press, 1982.

[Hellman, 1989] G. Hellman. *Mathematics without Numbers*, Oxford, Clarendon Press, 1989.

[Henkin and Montague, 1956] L. Henkin and R. Montague. On the Definitions of Formal Deduction, *The Journal of Symbolic Logic* **21**, 1956, pp. 129-136.

[Hilbert, 1902] D. Hilbert. Mathematical problems, *Bull. American Mathematical Society* **8**, 437-479, 1902, translated by M. W. Nelson from 'Mathematische probleme', *Archiv der Math. u. Phys.* **1**, 1901, 44-63 and 213-237. Reprinted in Browder 1976, pp. 1-34.

[Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*, Methuen, 1968.

[Jaśkowski, 1948] S. Jaśkowski. Rachuneck zdań dla systemów dedukcyjnych sprzecznych', *Studia Soc. Sci. Torunensis* **5**, pp. 55-77, 1948.

[Jaśkowski, 1949] S. Jaśkowski. O koniunkcji dyskusyjney w rachunku zdań dla systemów dedukcyjnych sprzecznych, *Studia Soc. Sci. Torunensis* **8**, pp. 171-172, 1949.

[Jaśkowski, 1969] S. Jaśkowski. Propositional calculus for contraditory deductive systems, *Studia Logica* **24**, pp. 143-157, 1969.

[Kaestner and Krause, 2996] C. A. A. Kaestner and D. Krause. Matrix proof in annotated paraconsistent logic, in *Atas del 2do. Workshop sobre Aspéctos Teóricos de la Inteligencia Artificial*, Universidad Nacional del Sur, Bahia Blanca, Argentina, 5-7 Octubre 1995, Univ. Nacional del Sur, pp. 416-425, 1996.

[Kifer and Kriphnaprasad, 1989] M. Kifer and T. Kriphnaprasad. An evidence based framework for a theory of inheritance, in *Proceedings of the 1989 International Joint Conference on Artificial Intelligence*, Morgan-Kauffman, San Mateo, CA, pp. 1093-1098, 1989.

[Kifer and Li, 1988] M. Kifer and A. and Li. On the semantics of rule-based expert systems with uncertainty, in *Proceedings of the 2nd. International Conference on Database Theory*, Bruges, Belgium, M. Gyssens, J. Paredaens and D. van Gucht (eds.), Springer Lecture Notes in Computer Science **326**, pp. 102-117, 1988.

[Kifer and Lozinskii, 1989] M. Kifer and E. L. Lozinskii. RI: A Logic for Reasoning with Inconsistency, *Proc. LICS*, pp. 253-262, 1989.

[Kifer and Subrahmanian, 1989] M. Kifer and V. S. Subrahmanian. Theory of generalized annotated logic programming and its applications, *Journal of Logic Programming* **12**, pp. 335-36, 1989.

[Kifer and Wu, 1989] M. Kifer and J. Wu. A logic for object oriented logic programming, in *Proceedings of the 8th. ACM SIGACT/SIGMOD/SIGART Symposium on Principlea of Database Systems*, Philadelphia, pp. 379-393, 1989.

[Kneale and Kneale, 1988] W. Kneale and M. Kneale. *The Development of Logic*, Oxford, Clarendon Press, 1988.

[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics*, New York, Van Nostrand, 1952.

[Kotas, 1971] J. Kotas. On the algebra of classes of formulas of Jaśkowski's discussive system, *Studia Logica* **27**, pp. 81-91, 1971.

[Kotas, 1974] J. Kotas. The axiomatization of Jaśkowski's discussive system, *Studia Logica* **28**, pp. 195-200, 1974.

[Kotas, 1975] J. Kotas. Discussive sentential calculus of Jaśkowski, *Studia Logica* **34**, 1975, pp. 149-168, 1975.

[Kotas and da Costa, 1977] J. Kotas and N. C. A. da Costa. On some modal logical systems defined in connection with Jaśkowski's problem. In Arruda, A. I., N. C. A. da Costa. A. and Chuaqui, R., (eds.), *Non-Classical Logic, Model Theory and Computability*, Amsterdam, North-Holland, pp. 57-73, 1977.

[Kotas and da Costa, 1978] J. Kotas and N. C. A. da Costa. On the problem of Jaśkowski and the logics of Łukasiewicz, in A. I. Arruda, N. C. A. da Costa and R. Chuaqui (eds.), *Mathematical Logic: Proceedings of the First Brazilian Conferece*, Marcel Dekker, pp. 127-139, 1978.

[Kotas and da Costa, 1979] J. Kotas and N. C. A. da Costa. A new formulation of discussive logic, *Studia Logica* **38**, pp. 429-445, 1979.

[Krause *et al.*, 2001] D. Krause, M. A. Musicante and E. F. Nobre. Bibel's matrix connection method in paraconsistent logic: general ideas and implementation, *Proceedings of the XXI Internacional Conference Chilean Computer Society*, Punta Arenas, Chile, 5-9 Nov. 2001 (IEEE Computer Society Press 2001), pp. 161-167, 2001.

[Kyburg, 1998] H. Kyburg. Review of [da Costa, 1974; da Costa, 1982; da Costa, 1986a; da Costa *et al.*, 1995], *The Journal of Symbolic Logic* **63** (3), pp. 1183-1184, 1998.

[Laugwitz, 1961a] D. Laugwitz. Anwendungen unendlichkliner Zahlen, I: Zur Theorie der Distributionen, *Journal f. d. reine und angewandte Mathematik* **207**, pp. 53-60, 1961.

[Laugwitz, 1961b] D. Laugwitz. Anwerdungen unendlichkliner Zahlen, II: Ein Zungang zur Operatorenrechnung von Mikusinski, *Journal f. d. reine und angewandte Mathematik* **208**, pp. 22-34, 1961.

[Lewin *et al.*, 1990] R. A. Lewin, I. F. Mikenberg and M. G. Schwarze. Algebrization of paraconsistent logic P_1, *The J. of Non-Classical Logic* **7** (1/2), pp. 145-154, 1990.

[Loparić, 1977] A. Loparić. Une étude sémantique de quelques calculs propositioneles, *C. R. Acad. Sc. Paris* **284A**, pp. 835-838, 1977.

[Loparić, 1978] A. Loparić. The method of valuation in modal logic, in A. I. Arruda, N. C. A. da Costa and R. Chuaqui (eds.), *Mathematical Logic: Proceedings of the First Brazilian Conferece*, Marcel Dekker, pp. 141-157, 1978.

[Loparić and Alves, 1980] A. Loparić and E. H. Alves. The semantics of the systems C_n of da Costa, in A. I. Arruda, N. C. A. da Costa and A. M. Sette (eds.), *Proceedings of the Third Brazilian Conference on Mathematical Logic*, Sociedade Brasileira de Lógica, São Paulo, pp. 161-172, 1980.

[Loparić and da Costa, 1986] A. Loparić and N. C. A. da Costa. Paraconsistency, paracompleteness, and induction, *Logique et Analyse* **113**, pp. 73-80. 1986.

[l'Hospital, 1996] G. F. A. l'Hospital. *Analyse des infinitament petits pour l'intelligence des lignes courbes*, Paris, 1696.

[Ludwig, 1990] G. Ludwig. *Les Structures de base d'une théorie physique*, Berlin, Springer-Verlag, 1990.

[Łukasiewicz, 1971] J. Łukasiewicz. On the principle of contradiction in Aristotle, *Review of Methaphysics* **24**, pp. 485-509, 1971.

[Manin, 1977] Yu. I. Manin. *A Course in Mathematical Logic*, Springer, 1977.

[Marconi, 1979] D. Marconi, ed. *La Formalizzazione della Dialettica*, Rosenberg & Sellier, Turin, 1979.

[Marconi, 1980] D. Marconi. A decision method for the calculus C_1, in A. I. Arruda, N. C. A. da Costa and A. M. Sette (eds.), *Proceedings of the Third Brazilian Conference on Mathematical Logic*, Sociedade Brasileira de Lógica, São Paulo, pp. 211-223, 1980.

[Marcos, 2004] Marcos, J., *Logics of Formal Inconsistency*, Ph.D. dissertation, Campinas, UNICAMP, 2004.

[Martins and Pequeno, 1993] A. T. C. Martins and T. Pequeno. Proof-theoretical considerations about the logic of epistemic inconsistency, *Logique et Analyse* **36**, no. 143-144, pp. 245-260, 1993.

[Martins and Pequeno, 1994] A. T. C. Martins and T. Pequeno. Proof-theoretical considerations about the logic of epistemic inconsistency, *Logique et Analyse*, 1996.

[Martins and Pequeno, 1996] A. T. C. Martins and T. Pequeno. A meta analyse of the inconsistent default logic reasoning style, submitted to the Special Issue of *The Journal of The Interest Group in Pure and Logics*, Oxford University Press, 1996.

[McGill and Parry, 1948] N. J. McGill and W. T. Parry. The unity of opposites: a dialectical principle, *Science and Society* **XII**, pp. 418-444, 1948.

[Mendelson, 1997] E. Mendelson. *Introduction to Mathematical Logic*, 4th. ed., London, Chapman & Hall, 1997.

[Mikenberg *et al.*, 1986] I. Mikenberg, N. C. A. da Costa and R. Chuaqui. Pragmatic truth and approximation to truth, *The Journal of Symbolic Logic* **51**, pp. 201-221, 1986.

[Milner, 1997] R. Milner, M. Tofte and R. Harper. *The Definition of Standard ML*, MIT Press, 1997.

[Moraes, 1970] L. de Moraes. *Sobre a Lógica Discursiva de Jaśkowski*, Master Thesis, University of São Paulo, S. Paulo, 1970.

[Mortensen, 1980] C. Mortensen. Every quotient algebra for C_1 is trivial, *Notre Dame J. of Formal Logic* **21**, pp. 694-700, 1980.

[Mortensen, 1995] C. Mortensen. *Inconsistent Mathematics*, Dordrecht, Kluwer Ac. Press, 1995.

[Nakamatsu *et al.*, 1999] K. Nakamatsu, J. M. Abe and A. and Suzuki. An approximate reasoning in a framework of vector annotated logic programming, *The Vietnam-Japan Bilateral Symposium on Fuzzy Systems And Applications*, VJFUZZY' 98, Nguyen H. Phuong & Ario Ohsato (Eds), HaLong Bay, Vietnam, pp. 521-528, 1998.

[Nakamatsu *et al.*, 2000] K. Nakamatsu, J. M. Abe and A. Suzuki. A railway interlocking safety verification systems based on abductive paraconsistent logic programming, *Soft Computing Systems: Design, Management and Applications*, Eds. A. Abraham, J. Ruiz-del-Solar & M. Köppen, Frontiers in Artificial Intelligence and Its Applications, IOS Press, Amsterdan, Ohmsha, Tokyo, Editores, Vol. 87, pp. 775-784, 2002.

[Nakamatsu *et al.*, 2001] K. Nakamatsu, J. M. Abe and A. Suzuki. 'Annotated semantics for defeasible deontic reasoning', in W. Ziarko and Y. Yao (eds.), *RCCTC 2000*, Springer Verlag, pp. 470-478, 2001.

[Nakamatsu *et al.*, 2002] K. Nakamatsu, J. M. Abe and A. Suzuki. 'Defeasible deontic robot control based on extended vector annotated logic programming', *Computer Antecipatory Systems: SASYS 2001 - Fifth International Conference*, ed. D. M. Dubois, American Institute of Physics, pp. 490-500, 2002.

[Nakamatsu *et al.*, to appear] K. Nakamatsu, H. Suito, J. M. Abe and A. Suzuki. Paraconsistent logic program based safety verification for air traffic control, to appear.

[Nakamatsu *et al.*, 2003] K. Nakamatsu, Y. Mita, T. Shibata and J. M. Abe. Defesiable deontic action control based on paraconsistent logic program and its implementation, *Proceedings of the 2003 International Conference on Computational Intelligence for Modelling, Contrl & Automation - CIMCA 2003*, M. Mohammadia (Ed), pp. 233-246, 2003.

[Negoita and Ralescu, 1975] C. V. Negoita and D. A. Ralescu. *Applications of fuzzy sets to systems analysis*, John Wiley & Sons, New York, 1975.

[Nelson, 1949] D. Nelson. Constructive falsity', *J. Symb. Logic* **14**, pp. 16-26, 1949.

[Nelson, 1959] D. Nelson. Negation and separation of concepts in constructive mathematics, in Heyting, A. (ed.), *Constructive Mathematics*, North-Holland, pp. 208-225, 1959.

[Peirce, 1965] C. S. Peirce. *Philosophical Writings of Peirce*, selected and edited by Buchler, J., New York, Dover, 1965.

[Peña, 1979] L. Peña. *Contradiction et Verité*, Thesis, Université de Liège, 1979.

[Peña, 1991] L. Peña. *Rudimentos de Lógica Matemática*, Consejo Superior de Investigaciones Científicas, Collectión Textos Universitarios 7, Madrid, 1991.

[Peña, 1993] L. Peña. *Introducción a las Lógicas No Clásicas*, Instituto de Investigaciones Filosóficas, Cuadernos 60, Universidad Nacional Autónoma de México, 1993.

[Pequeno, 1990] T. Pequeno. A logic for inconsistent non-monotonic reasoning, *Technical Report* **90/6**, Department of Computing, Imperial College, London, 1990.

[Pequeno, 1991] T. Pequeno and A. Buchsbaum. The logic of epistemic inconsistency, *Principles of knowledge representation and reasoning*, Cambridge, MA, pp. 453-460, 1991.

[Pequeno and Buchsbaum, 1995] T. Pequeno and A. Buchsbaum. The logic of epistemic inconsistency, in Carnielli, W. A. and Pereira, L. C. P. D., *Logic, sets and information*, Proccedings of the Tenth Brazilian Conference on Mathematical Logic, Universidade Estadual de Campinas, Coleção CLE **14**, pp. 177-197, 1995.

[Pinter, 1980] C. Pinter. The logic of ambiguity', in A. I. Arruda, N. C. A. da Costa and A. M. Sette (eds.), *Proceedings of the Third Brazilian Conference on Mathematical Logic*, Sociedade Brasileira de Lógica, São Paulo, pp. 253-262, 1980.

[Pollock, 1987] J. L. Pollock. Defeasible reasoning, *Cognitive Science*, **11**, pp. 481-518, 1987.

[Popper, 1972] K. R. Popper. *The Logic of Scientific Discovery*, Hutchinson, 1972,

[Priest, 1979] G. Priest. The logic of paradox, *The Journal of Philosophical Logic* **8**, pp. 219-241, 1979.

[Priest, 1987] G. Priest. *In Contradiction*, Dordrecht: Nijhoff, 1987.

[Priest *et al.*, 1989] G. Priest, R. Routley and J. Norman, eds. *Paraconsistent Logic: essays on the inconsistent*, Philosophia Verlag, Munich, 1989.

[Priest and Tanaka, 2000] G. Priest and K. Tanaka. Paraconsistent logic, *The Stanford Encyclopedia of Philosophy* (Winter 2000 Edition), Edward N. Zalta (ed.), URL = ¡http://plato.stanford.edu/archives/win2000/entries/logic-paraconsistent/¿.

[Puga, 1985] L. Z. Puga. *Uma Lógica do Querer: preliminares sobre um tema de Mally*, Doctor Thesis, Ponthifical Catholic University of São Paulo, S. Paulo, 1985.

[Puga and da Costa, 1987] L. Z. Puga and N. C. A. da Costa. Logic with deontic and legal modalities, *Bull. Sec. Logic of the Polish Academy of Sciences* **2**, pp. 141ss, 1987.

[Puga and da Costa, 1988] L. Z. Puga and N. C. A. da Costa. On the imaginary logic of N. A. Vasiliev, *Zeitschr. f. math, Logik und Grundlagen d. Math.* **34**, pp. 205-211, 1988.

[Puga and da Costa, 1992] L. Z. Puga, N. C. A. da Costa, and R. J. Vernengo. Normative logics, morality and law, *Expert Systems in Law*, A. Martino (ed.), Elsevier Sci. Pu., , pp. 345-365, 1992.

[Robinson, 1966] A. Robinson. *Non-standard analysis*, North-Holland, 1966.

[Rosser, 1953] B. Rosser. *Logic for Mathematicians*, New York, McGraw-Hill, 1953.

[Routley and Loparić, 1980] R. Routley and A. Loparić. Semantics for quantified relevant logics without replacement, in A. I. Arruda, N. C. A. da Costa and A. M. Sette (eds.), *Proceedings of the Third Brazilian Conference on Mathematical Logic*, Sociedade Brasileira de Lógica, São Paulo, pp. 262-280, 1980.

[Routley, 1979] R. Routley. Dialectical logic, semantics and metamathematics, *Erkenntnis* **14**, pp. 310-331, 1979.

[Routley and Loparić, 1979] R. Routley and A. Loparić. Semantical analysis of Arruda and da Costa *P* systems and adjacent non-replacement systems, *Studia Logica* **37**, pp. 301-320, 1979.

[Routley and Meyer, 1976] R. Routley and R. K. Meyer. Dialectical logic, classical logic and the consistency of the world, *Studies in Soviet Thought* **16**, pp. 1-25, 1976.

[Scott, 1967] D. Scott. A proof of the independence of the contiuum hypothesis, *Mathematical Systems Theory* **1**, pp. 89-111, 1967.

[Schmieden and Laugwitz, 1958] C. Schmieden and D. Laugwitz. Eine Erweiterung der Infinitesimalrechnung, *Math. Zeitschrift* **69**, pp. 1-39, 1958.

[Sette, 1971] A. M. Sette. *Sobre as Álgebras e Hiper-Reticulados* C_n , Master Thesis, Universidade Estadual de Campinas, 1971.

[Shoenfield, 1967] J. R. Shoenfield. *Mathematical Logic*, Reading, Addison-Wesley, 1967 (reprinted by the Association of Symbolic Logic, 2000).

[Silva Filho and Abe, 2001a] J. I da Silva Filho and J. M. Abe. Paraconsistent electronic circuits, *International Journal of Computing Anticipatory Systems*, **9**, pp. 337-345, 2001.

[Silva Filho and Abe, 2001b] J. I. da Silva Filho and J. M. Abe. Emmy: a paraconsistent autonomous mobile robot, in *Logic, Artificial Intelligence, and Robotics*, Proc. 2nd Congress of Logic Applied to Technology - LAPTEC 2001, Eds. J.M. Abe & J.I. Da Silva Filho, Frontiers in Artificial Intelligence and Its Applications, IOS Press, Amsterdan, Ohmsha, Tokyo, Editores, Vol. 71, pp. 53-61, 2001.

[Subrahmanian, 1987] V. S. Subrahmanian. On the semantics of quantitative logic programs, Proc. 4th IEEE Symposium on Logic Programming, Computer Socitety Press, Washington, DC, pp. 173-182, 1987.

[Tuziak, 1993] R. Tuziak. Review of Priest, Routley's and Norman (eds.) 'Paraconsistent logic: essays on the inconsistent', *Brit. J. Phil. Sci.* **44**, pp. 167-170, 1993.

[Ullian, 1990] J. S. Ullian. Learning and meaning, in Barrett, R. B. and Gibson, R. F., (eds.), *Perspectives on Quine*, Oxford, Blackwell, pp. 336-346, 1990.

[Urbas, 1989] I. Urbas. Paraconsistency and the C_1-systems of da Costa, *Notre Dame J. of Formal Logic* **30** (4), pp. 583-597, 1989.

[Vaihinger, 1952] H. Vaihinger. *Philosophy of 'As If': A System of Theoretical, Practical and Religious Fictions of Mankind*, London, Routledge & Kegan Paul, 1952.

[van Fraassen, 1980] B. van Fraassen. *The Scientific Image*, Oxford, Clarendon Press, 1980.

[Yamashita, 1985] M. Yamashita. *O símbolo de Hilbert em Lógica Paraconsistente*, Doctor Thesis, Ponthifical Catholic University of São Paulo, S. Paulo, 1985.

[Zadeh, 1978] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility, *Fuzzy sets and Systems*, **1**, pp. 3-28, 1978.

[Zhang, 2003] Q. Zhang. *Paraconsistent Logic* [in Chinese], Chinese Society Press, 2003.

EXTENSIONAL VS. INTENSIONAL LOGIC

Jaroslav Peregrin

1 FREGE'S *BEDEUTUNG*

The German mathematician and one of the founding fathers of modern logic, Gottlob Frege (1848–1925), was the first to clearly realize that semantics has little to do with psychology, and that it could be usefully explicated in mathematical terms (see [Dummett, 1973; 1981]). His depsychologization of semantics followed from his depsychologization of logic. Frege understood how crucial it was for the development of logic to draw a sharp boundary separating it from psychology: to make it clear that logic is not a matter of what is going on in some person's head, in the sense that psychology is. This is because logic is concerned with what is true and consequently what follows from what — and whether something is true, or whether something follows from something else, is an objective matter independent of what is going on in the head of a particular person.

As a consequence, Frege realized that if logic must be separated from psychology, then the same is true for semantics — at least insofar as semantics underlies truth and entailment. It is clear that the truth value of a sentence depends on the meaning of the sentence: the sentence "Penguins eat fish" is true not only due to the fact that penguins do eat fish, but of course also due to the fact that the words of which it consists mean what they do in English. Hence, if meaning were a matter of what is going on in somebody's head, then truth would have to be too — hence meaning must not, in pain of the subjectivization of truth, be a psychological matter. But what, then, *is* meaning?

Frege started from the *prima facie* obvious fact that names stand for objects of the world. Unprecedentedly, he assimilated indicative sentences to names as well: he saw them as specific kinds of names denoting the two truth values: *truth* and *falsity*. The reason for this move was that he divided expressions into two sharply separated groups: into "saturated" — i.e. self-standing — and "unsaturated" — i.e. incomplete — ones. He took names and sentences as species of the former kind, whereas he took predicates as paradigmatic examples of the latter one; and he came to use the word "name" as a synonym of "saturated expression".

The reason why he identified the entities named by sentences with truth values was articulated by Frege in the form of what has later become known as the *slingshot argument*.[1] This argument itself rests on what has subsequently come to be called *the principle of compositionality* and what Frege tacitly, but unambiguously assumed.[2] The principle

---

[1] See [Neale, 2001].

[2] See [Janssen, 1997].

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

states that the meaning of a complex expression is uniquely determined by the meanings of its parts plus the mode of their combination. This means that for every mode of combination (grammatical rule) G there must exist a function G^* such that for every $e_1, ..., e_n$ to which G is applicable it is the case that (where $||e||$ denotes the meaning of e):[3]

$$||G(e_1, ..., e_n)|| = G^*(||e_1||, ..., ||e_n||).$$

It follows that replacing a part e of a complex expression by an expression e' with the same meaning as e we cannot change the meaning of the whole complex:

$$\text{if } ||e|| = ||e'||, \text{ then } ||G(..., e, ...)|| = ||G(..., e', ...)||.$$

Now Frege argued that replacing names by other names of the same entities can get us from a sentence to a sentence which has very little in common with the original one; in particular that there is only one thing which inevitably persists during such a process, and this is precisely the truth value.

We can illustrate this by means of an example introduced by Alonzo [Church, 1956, 24–25]:

> [T]he denotation (in English) of "Sir Walter Scott is the author of *Waverley*" must be the same as that of "Sir Walter Scott is Sir Walter Scott," the name "the author of *Waverley*" being replaced by another which has the same denotation. Again, the sentence "Sir Walter Scott is the author of *Waverley*" must have the same denotation as the sentence "Sir Walter Scott is the man who wrote twenty-nine Waverley Novels altogether," since the name "the author of *Waverley*" is replaced by another name of the same person; the latter sentence, it is plausible to suppose, if it is not synonymous with "The number, such that Sir Walter Scott is the man who wrote that many Waverley Novels altogether, is twenty-nine," is at least so nearly so as to ensure its having the same denotation; and from this last sentence in turn, replacing the complete subject by another name of the same number, we obtain, as still having the same denotation, the sentence "The number of counties in Utah is twenty-nine."
>
> Now the two sentences, "Sir Walter Scott is the author of *Waverley*" and "The number of counties in Utah is twenty-nine," though they have the same denotation according to the preceding line of reasoning, seem actually to have very little in common. The most striking thing that they do have in common is that both are true. Elaboration of examples of this kind leads us quickly to the conclusion, as at least plausible, that all true sentences have the same denotation. And parallel examples may be used in the same way to suggest that all false sentences have the same denotation (e.g., "Sir Walter Scott is not the author of *Waverley*" must have the same denotation as "Sir Walter Scott is not Sir Walter Scott").

---

[3] In algebraic terms, meaning-assignment is a *homomorphism*.

Hence it seems that if we accept that a name means the thing it names, then given the principle of compositionality, two sentences are bound to have the same meaning iff they have the same truth value; and it is plausible to explicate the meanng of a sentence with its truth value.

Frege's most brilliant contribution to the explication of the concept of meaning then was the way he accounted for the meanings of predicates.[4] He called them *concepts*, as usual; but he rejected the standard view of concepts as something mental and, in effect, suggested explicating them by studying the role of those expressions which express them — i.e. predicates — within language.

What is the role of a predicate, such as "to sing"? Well, the predicate is attached to a subject, a name such as "Frege", to form a sentence, "Frege sings". Hence if we assume that the meaning of a complex expression is the result of combining the meanings of its parts (i.e. that meanings are composed in a way paralleling that in which the expressions expressing them are), then the meaning of the predicate together with the meaning of a subject, which is the object stood for by the subject, yields the meaning of a sentence, i.e. a truth value. Therefore a concept is something that together with an object yields a truth value — and this led Frege to identify concepts with functions, in the mathematical sense of the word, taking objects to truth values. In effect, this meant the identification of the meaning of an item with the semantic role of the item captured as a function in the mathematical sense of the word; and this opened the door for a mathematical treatment of semantics. Thus, we can say that Frege married semantics, which he had earlier divorced from psychology, to mathematics.

Let us reconstruct this move, which we will dub *Frege's maneuver*, in greater detail. We have a category of expressions, call it A, whose meanings we want to explicate. We see that an expression of the category A (say that of predicates) can be combined with an expression of a category B (say that of names, singular terms) to form an expression of a category C (say that of sentences). So if we take an expression a of the category A, we know that together with the expression b_1 of the category B it yields an expression c_1 of C, with b_2 it yields c_2 etc.:

$$\begin{array}{ll} a + b_1 = c_1 & \text{“to sing” + “Madonna” = “Madonna sings”} \\ a + b_2 = c_2 & \text{“to sing” + “Moon” = “Moon sings”} \\ \dots & \dots \end{array}$$

Hence we can see a as a means of assigning c_1 to b_1, c_2 to b_2 etc.

$$\begin{array}{rlrl} a: & b_1 \to c_1 & \text{“to sing”:} & \text{“Madonna”} \to \text{“Madonna sings”} \\ & b_2 \to c_2 & & \text{“Moon”} \to \text{“Moon sings”} \\ & \dots & & \dots \end{array}$$

Now suppose that we know what the meanings of both the expressions of B and of C are. Then we can transfer the whole consideration to the level of semantics:

$$\begin{array}{rlrl} ||a||: & ||b_1|| \to ||c_1|| & ||\text{to sing}||: & ||\text{Madonna}|| \to ||\text{Madonna sings}|| \\ & ||b_2|| \to ||c_2|| & & ||\text{Moon}|| \to ||\text{Moon sings}|| \\ & \dots & & \dots \end{array}$$

---

[4] See especially [Frege, 1891; 1892a].

Now if we accept that meanings of names are the objects named by them (the meaning of "Madonna", ||Madonna||, is the person Madonna, that of "Moon", ||Moon||, is the celestial body the Moon) and that the meanings of sentences are their truth values (||Madonna sings|| being *the truth*, **T**, that of ||Moon sings|| being *the falsity*, **F**), what we gain by this maneuver as the explication of the meaning of "to sing" is a function assigning truth values to individuals: **T** to those which do sing (to Madonna and others) and **F** to those which do not (to the Moon and others).

Let us further illustrate this maneuver by applying it to logical operators. Take conjunction: it joins two sentences to form a sentence. Thus it can be considered as assigning sentences to pairs of sentences. Hence on the semantic level it can be seen as assigning truth values to pairs of truth values; and hence its meaning can be identified with a function taking pairs of truth values to a truth value. Which particular function? This is revealed by inspecting the dependence of the truth values of conjoined sentences on those of their parts; and of course it is the usual well-known function assigning **T** to **T** plus **T**, and **F** to any other pair of truth values. Thus, according to Frege, meanings were either objects or functions, and the principal way of combining meanings of parts of the whole expression into the meaning of the whole was functional application.

Notice that Frege's maneuver has two substantial presuppositions. First, there is the presupposition that the meaning of a complex expression is yielded by (or 'composed of') meanings of the parts of the expressions — *viz.* the principle of compositionality. How do we know that this principle holds? Some theoreticians appear to think that it is an empirical thesis that must be verified as empirical theses are: by means of inspecting as many cases as possible. However, such a view presupposes that meanings are independently identifiable objects whose combinations can be studied in the way we study, e.g., combinations of molecules in a solution: that we can empirically verify (or falsify) the thesis that, say, the meaning of a sentence is yielded by the meaning of its subject and that of its predicate, by means of finding the meanings and finding out what they yield if they are put together. In contrast to this, we saw that for Frege the principle was rather a way of articulating what it takes to be meaning: the principle was co-constitutive of the notion of meaning in a sense analogous to that in which, say, the principle of extensionality is co-constitutive of the concept of set. And just as it makes no sense to try to find out whether sets are extensional (for this is simply part of what it takes to be a set), it makes no sense to try to find whether meaning is compositional.[5]

The other presupposition of Frege's maneuver is of a different kind: it concerns the behavior of the particular expressions to which it is applied. The presupposition is that the role of the expression within language is exhausted by, or at least in some sense reducible to, its role within the kind of syntactic combination which is taken into consideration. We explicated the meanings of predicates by considering the way they combine with names into sentences; but predicates also do other things, e.g. combine with adverbials into complex predicates. We must always be sure that this is taken care of — that it is proven that it is somehow substantiated to treat some part of the functioning of an expression as representative of the whole functioning.

---

[5] See [Peregrin, 2005b]. For general discussions of the principle of compositionality see also [Janssen, 1997; 2001] and [Pelletier, 2001].

Is Frege's way of explicating the concept of meaning acceptable? In fact not: what Frege called *meaning* cannot be taken as a plausible explication of the pre-theoretic notion of meaning. After all, who would want to claim that all true sentences have the same *meaning*? And Frege himself soon realized the implausibility of such an explication. Therefore he complemented his theory of meaning with what he called a theory of *sense*.[6] Every name, he claimed, has not only a meaning, but rather also a sense, which is the 'way of givenness' of the meaning. And it is then Frege's concept of *sense*, rather than his concept of *meaning*, which is to be taken as his explication of the intuitive concept of meaning.

Frege's own instructive example is that of the terms "morning star" and "evening star". As we now know, these two terms refer to one and the same celestial body, the planet Venus. Hence they share the same meaning (in Frege's sense of the word), or (in the current jargon) the same *referent*. If the referent were all that there is to meaning, it should be possible to substitute one of them for the other within any expression without changing its meaning. However, although the sentence "The morning star is the morning star" is obviously trivial, "The morning star is the evening star" does not appear to be such. The reason, Frege claimed, is that the terms differ in their senses, i.e. in the ways they present their referent: "the morning star" presents it as the most attractive body in the morning sky, whereas "the evening star" as the most attractive one in the evening sky.

Hence we have the general picture according to which the relation between a name and what the name refers to is mediated by the sense of the name:

NAME
↓
SENSE (*SINN*)
(= meaning in the intuitive sense of the word)
↓
MEANING (*BEDEUTUNG*)
(= object referred to)

One must remember, though, that despite his recognition of *Sinn*, Frege kept insisting that it is *Bedeutung* which is crucial from the viewpoint of logic. As he puts it in an unpublished text [1983, 133]:

> Die Inhaltslogiker bleiben nur zu gerne beim Sinn stehen; denn, was sie Inhalt nennen, ist, wenn nicht gar Vorstellung, so doch Sinn. Sie bedenken nicht, dass es in der Logik nicht darauf ankommt, wie Gedanken aus Gedanken hervorgehen ohne Rücksicht auf den Wahrheitswert, dass, allgemeiner, der Schritt vom Sinne zur Bedeutung getan werden muss; dass die logischen Gesetze zunächst Gesetze im Reich der Bedeutungen sind und sich erst mittlebar auf denn Sinn beziehen.[7]

---

[6] See [Frege, 1892b].

[7] "The content-logicians only remain too happily with the sense, for what they call content is for them, if not mental image, then surely sense. They do not consider the fact that in logic it is not a question of how thoughts

To summarize: Frege saw as his main target *truth* (and *truth-dependence*, i.e. consequence); and took for granted that the account for it is to be *compositional*. He concluded that this can be accomplished on the level of his *meaning* (*Bedeutung*); in particular that (i) the assignment of *meanings* to expressions is compositional; (ii) the *meaning* of a sentence *is* its truth value. (The most developed logical system based on the ideas, presented in his *Grundgesetze der Arithmetik*,[8] was, however, shown to be inconsistent.)

2 CARNAP'S *EXTENSION* AND EXTENSIONAL LOGIC

Rudolf Carnap (1891–1970), the logician and logico-positivist philosopher, realized that if what we are after is meaning in the intuitive sense of the word, then we should be interested not so much in meanings in the sense of Frege, but rather in Fregean *senses*. However, as Frege did not explicate the concept of sense to Carnap's satisfaction, Carnap proposed replacing the Fregean twin concepts of *meaning* and *sense* with the concepts of *extension* and *intension*. (A laudable proposal, for it did away with Frege's misleading usage of "meaning".) And his claim is that the *extension* of a term is what the term shares with all terms that are equivalent to it; whereas its intension is what it shares with all the terms that are *logically* equivalent to it.

Of course this definition becomes nontrivial only after we give a rigorous account of the concept of equivalence on which it rests. For the basic categories of the predicate calculus this is not difficult: two individual terms t_1 and t_2 are equivalent iff

$$t_1 = t_2,$$

two n-ary predicates p_1 and p_2 are equivalent iff

$$\forall x_1 \dots \forall x_n (p_1(x_1, \dots, x_n) \leftrightarrow p_2(x_1, \dots, x_n));$$

and two sentences s_1 and s_2 are equivalent iff

$$s_1 \leftrightarrow s_2.$$

This explication leads to a concept of extension almost indistinguishable from Frege's *meaning*: the extension of an individual expression being the object for which it stands, that of a predicate being the function assigning the truth value **T** to those n-tuples of objects of which the predicate is true, and that of a sentence being its truth value.

Meanwhile, what was built into the foundations of modern formal logic were extensions. The basic concept of a *semantic interpretation*, due essentially to Tarski and accepted by the mainstream of logicians,[9] was in fact that of the mapping of expressions on their extensions: names on individuals, predicates on sets of individuals or sets on their

---

come from thoughts without regard of truth value, that, more generally, one must make the step from senses to meanings; that the logical laws are first laws in the realm of meanings and only then mediately relate to senses."

[8] [Frege, 1893; 1903].

[9] The roots of the concept go back to Tarski [1936] (cf. also [Etchemendy, 1990]). For the definition of the concept of semantic interpretation in the modern sense, see, e.g. [Mendelson, 1964].

n-tuples[10] and sentences on truth values. The reasons why most logicians did not share Carnap's scruples were, roughly, two: (i) like Frege, they assumed that what is the primary aim of logic is *truth*, and that what we need to account for truth were extensions, not intensions; (ii) many of them were primarily interested in mathematics, where intensions, explicated in the Carnapian way, came to coincide with extensions. Many logicians of the second half of the twentieth century, especially those inclining to mathematics, thus came to the conclusion that there is something as *the* language of logic — the language of first-order predicate calculus (FOPC). And the semantics of this language is based on mapping expressions on their extensions.

To see this, consider FOPC as it is standardly presented. We have the logical constants $\neg$, $\wedge$, and $\forall$; the extralogical (individual, predicate, functor) constants and the (individual) variables, with the well-known syntactic rules. The semantics of the language is then based on two sets of assignments of objects to expressions: the *valuation of variables* (assignments of individuals to variables) and the *interpretation of extralogical words* (assignments of individuals to individual constants, sets of n-tuples of individuals to n-ary predicates and, n-ary functions from individuals to individuals to functor constants). Given an interpretation I and a valuation V, every individual term t is assigned a denotation $||t||_{I,V}$ based on I and V so that $||t||_{I,V} = I(t)$, if t is an individual constant, $||t||_{I,V} = V(t)$ if t is an individual variable and $||f(t_1,\ldots,t_n)||_{I,V} = I(f)(||t_1||_{I,V},\ldots,||t_n||_{I,V})$. An interpretation I and a valuation V then render each formula true or false, which is defined in the usual recursive way: $p(t_1,\ldots,t_n)$ is true w.r.t. (or satisfied by) I and V iff $\langle ||t_1||_{I,V},\ldots,||t_n||_{I,V}\rangle \in I(p)$; $\neg F$ is true w.r.t. (or satisfied by) I and V iff F is not; $F_1 \wedge F_2$ is true w.r.t. (or satisfied by) I and V iff both F_1 and F_2 are; $\forall xF$ is true w.r.t. (or satisfied by) I and V iff F is true w.r.t. (or satisfied by) I and V' for every valuation V' which differs from V at most in the value it assigns to x.

In this way, only extralogical words are taken to denote extensions; whereas the logical ones are treated as nondenoting terms. However, it is easy to do away with this difference in status by redefining an interpretation as a mapping of *all* expressions on their extensions, reducing the difference between logical and extralogical expressions to the mere fact that whereas the former ones have *fixed* interpretation, the interpretation of the latter *varies*. We may let logical constants denote the usual truth functions and we can let quantifiers denote functions mapping functions from individuals to truth values on truth values; in particular $\exists$ as denoting the function which maps f on **T** iff f maps at least one individual on **T** and $\forall$ as denoting the function which maps f on **T** iff f maps all individuals on **T**. If we do this, we can reformulate the above definition of truth as an explicit definition of denotation-assignment:

$$||p(t_1,\ldots,t_n)||_{I,V} = ||p||(||t_1||_{I,V},\ldots,||t_n||_{I,V})$$
$$||\neg F||_{I,V} = ||\neg||(||F||_{I,V})$$
$$||F_1 \wedge F_2||_{I,V} = ||\wedge||(||F_1||_{I,V}, ||F_2||_{I,V})$$
$||\forall xF'||_{I,V} = ||\forall||(f)$, where f is the function which maps an individual i on $||F||_{I,V'}$, where V' is just like V with the single possible exception that it maps x on i.

---

[10] There is usually no need to make a difference between a set and its *characteristic function*, i.e. the function which assigns **T** to elements of the set and **F** to elements of its complement.

In this way we can see semantic interpretation, i.e. the function $||\ldots||$, as simply an assignment of extensions.

Notice that viewed from this perspective denotations of all expressions of the language can be seen as arising from the denotations of terms and sentences by means of Frege's maneuver. Thus, a predicate takes an n-tuple of terms into a sentence, and hence it denotes an n-ary function from individuals to truth values; a functor takes an n-tuple of terms into a term, hence it denotes an n-ary function from individuals to individuals; and a sentential operator takes one or two sentences to a sentence, hence denotes a function from truth values or their pairs to truth values. The situation is trickier w.r.t. quantifiers, which *prima facie* do *not* take predicates to sentences (which would seem to have to underlie the application of Frege's maneuver resulting into the above kind of denotations for quantifiers).

However, to straighten this it is only necessary to reassess the underlying syntax. Let us introduce a new kind of rule, the rule of *lambda-abstraction*,[11] taking a formula and a variable (typically one contained free in the formula) to a unary predicate, denoting the function which maps an individual i of the universe on the extension which the formula would have if the variable denoted i. Formally, if F is a sentence and x a variable, then $\lambda x.F$ is a unary predicate whose semantics is defined in the following way:

> $||\lambda x.F||_{I,V}$ is the function which maps an individual i on $||F||_{I,V'}$, where V' is just like V with the single possible exception that it maps x on i.

Given this, we can treat a quantifier as an expression taking a unary predicate to a sentence, and take the quantified sentence $Qx.F$ as a shortcut for $Q(\lambda x.F)$ so that

$$||\forall x.F||_{I,V} = ||\forall(\lambda x.F)||_{I,V} = ||\forall||(||\lambda x.F||_{I,V}); \text{ and}$$

$$||\exists x.F||_{I,V} = ||\exists(\lambda x.F)||_{I,V} = ||\exists||(||\lambda x.F||_{I,V}).$$

This reformulation of FOPC renders all standard kinds of syntactic rules as paralleled by the same kind of combination of denotations — namely *functional application*: it is always the application of one of the denotations to the rest. (We have seen that this is the result of the fact that with the exception of terms and sentences, we can see the denotations of all other categories as furnished via Frege's maneuver.) Hence we can see the usual kind of extensional semantics as based on:

(i) letting terms denote the objects they name;

(ii) letting sentences denote truth values (which derives from (i) via the slingshot); and

(iii) letting expressions of all other categories stand for functions constructed from individuals and truth values (which derives from (i) and (ii) via Frege's maneuver).

Probably the most general language of extensional logic, based on the most consequential exploitation of Frege's maneuver, was put forward by Alonzo [Church, 1940] under the name of the *simple theory of types*.[12] Church's language is based on two basic categories, the category ι of individual terms and the category o of sentences. Besides this, there is a category $(\alpha\beta)$ for every two categories α and β of the language. (Thus we have categories such as (oo), $(o\iota)$, $(o(o\iota))$ etc.) The extensional semantics of the language is such that expressions of the category ι denote individuals (elements of a given universe of discourse), expressions of the category o denote truth values, and expressions of every category $(\alpha\beta)$ denote functions from the denotations of the expressions of β to those of the expressions of α. Thus, expressions of the category (oo) denote functions from truth values to truth values, those of $(o\iota)$ functions from individuals to truth values, those of $(o(o\iota))$ functions from these functions to truth values etc.

---

[11] See [Church, 1956].

Given this semantics, the basic kind of grammatical rule of the language forms the expression $a(b)$ of the category α from an expression a of the category $(\alpha\beta)$ and an expression b of the category β; and the denotation of the resulting expression $a(b)$ results from the application of the denotation of a to that of b. To this, Church only added the above rule of lambda abstraction. The resulting language was syntactically quite rich, but wonderfully simple:

Church's Simple theory of types (Ty1)

Syntax

o and ι are types, and if α and β are types then also $(\alpha\beta)$ is a type. For each type there is an unlimited stock of constants and variables of the type.[13] If a is an expression of type $(\alpha\beta)$ and b an expression of type β, then $a(b)$ is an expression of type α; and if a is an expression of type α and x a variable of type β, then $\lambda x.a$ is an expression of type $(\alpha\beta)$.

Semantics

To each type α there corresponds a domain D_α, so that D_o is the set of the two truth values, D_ι is a universe of discourse (any given set) and $D_{(\alpha\beta)}$ is a set of functions from D_β to D_α. An interpretation maps every constant a of type α on an element of D_α; a valuation does the same for variables. If I is an interpretation and V a valuation we define the denotation assignment $||\ldots||_{I,V}$ as follows: for a constant c, $||c||_{I,V} = I(c)$; for a variable v, $||v||_{I,V} = V(v)$. Moreover, $||a(b)||_{I,V} = ||a||_{I,V}(||b||_{I,V})$ and $||\lambda x.a||_{I,V}$ is the function which maps every element i of D_β on $||a||_{I,V'}$, where V' is just like V with the single possible exception that it maps x on i.

---

[12] The generalization leading to this language is parallel to the one proposed by Ajdukiewicz (1935) and [Bar-Hillel, 1950] and which has led to what is now called *categorial grammar* (see [Casadio, 1988]). [Bar-Hillel, 1953, 65] expounds the idea behind this as follows: "Each sentence which is not an element is regarded as the outcome of the operation of one sub-sequence upon the remainder, which may be to its immediate right, or to its immediate left or on both sides. ('Left' and 'right' are to be understood here, as in what follows, only as the two directions of a linear order.) That sub-sequence which is regarded as operating upon the others will be called an *operator*, the others its *arguments*."

[13] The construals of languages of this kind often fluctuate between the delimitation of *specific* (though general) *language* with a maximal vocabulary so that more specific languages can be cast as its sublanguages; and the delimitation of a mere *language form* with vocabulary 'open' and to be specified only on the level of specific languages.

It is not difficult to embed the language of FOPC into this language: n-ary predicates become expressions of the category $(((o\iota)\iota)\ldots\iota)$, n-ary terms those of $(((\iota\iota)\iota)\ldots\iota)$,[14] propositional operators turn into expressions of the category (oo) or $((oo)o)$, whereas quantifiers become expressions of the category $(o(o\iota))$. The syntactical structure of Church's language is, however, incomparably richer than that of the language of FOPC.

Consider an axiomatization. We start from some axiom system of the classical propositional calculus, say

$$\begin{array}{l} S_1 \to (S_2 \to S_1) \\ (S_1 \to (S_2 \to S_3)) \to ((S_1 \to S_2) \to (S_1 \to S_3)) \\ (\neg S_1 \to \neg S_2) \to (S_2 \to S_1) \\ S_1, S_1 \to S_2 / S_2 \end{array}$$

(where S_1, S_2 and S_3 are expressions of the type o, $\neg$ and $\to$ are constants of the respective types (oo) and $((oo)o)$; and we write $\neg S$ and $S \to S'$ instead of $\neg(S)$ and $(\to (S))(S')$, respectively. We add a generalization of the usual additional axioms and rule of the classical first-order predicate calculus — for every type α we have:

$$\begin{array}{l} \Pi_\alpha(F) \to F(A) \\ \Pi_\alpha(\lambda x.(S \to F(x))) \to (S \to \Pi_\alpha(F)), \text{ where } x \text{ is not free in } S \\ S / \Pi_\alpha(\lambda x.S) \end{array}$$

(Here S is an expression of the type o, F is an expression a type $(o\alpha)$, A is an expression of type α, x is a variable of type α, and Π_α is a constant of the type $(o(o\alpha))$.) Now it turns out that if we define $=_\alpha$ as $\lambda a.\lambda b.\Pi_o(\lambda f.(f(a) \to f(b)))$ (where f is a variable of type $(o\alpha)$), and $a$and b those of type α) all that must be added to the above axioms to obtain an axiomatization of Ty1 are the following four axiom schemes (again, for every type α and writing the equality sign in the obvious infix way):

$$\begin{array}{l} (S_1 \leftrightarrow S_2) \to (S_1 =_o S_2) \\ \Pi_\alpha(\lambda x.(F(x) =_o G(x))) \to (F =_{(o\alpha)} G) \\ (\lambda x.B)(A) =_\beta B^{x \leftarrow A} \\ F(A) \to F(\iota_\alpha(F)) \end{array}$$

(where S_1, S_2, F, x are as above, ι_α a constant of type $(\alpha(o\alpha))$, G an expression of type $(o\alpha)$, B an expression of type β, $S \leftrightarrow S$' is a shorthand for $\neg((S \to S') \to \neg(S' \to S))$, and $B^{x \leftarrow A}$ denotes the result of the replacement of x by A throughout B).

Note that as $A =_\alpha B$ amounts to A and B being intersubstitutive *salva veritate*, the first two of the axioms guarantee the 'extensionality' of the system: they guarantee that equivalent statements are always equal (hence that sentences can be treated as denoting truth values) and that functions which have the same courses-of-values are equal (and hence that they can be treated as denoting the very courses-of-values, which is the essence of Frege's maneuver). The third axiom, called *lambda-conversion*, guarantees the proper semantic working of the λ-operator;[15] whereas the last axiom is a sort of an axiom of choice that guarantees that whenever a set is not empty, we can name an element of the set (which implies that whenever there is a true existential statement, we have the name of a corresponding witness).[16]

---

[14] In this way, we take an n-ary predicate (functor) to take not an n-tuple of terms to a sentence (term), but rather a term to an (n-1)-ary predicate (functor); hence the combination of the n-ary predicate (functor) with n individuals comes out as n subsequent applications. This purely formal maneuver makes it possible to make do with merely unary predicates (functors).

Completeness of a slightly different version of this axiom system was proved by Henkin [1950]. However, the completeness proof is possible only because we defined the semantics so that $D_{(\alpha\beta)}$ is *a* set of functions from D_β to D_α; once we replace "a set of functions" by "*the* set of *all* functions", we get a semantics which is not axiomatizable (in consequence of the Gödel theorem). The former kind of semantics is usually called the *Henkin semantics*, whereas the latter is called *standard semantics*.[17] Hence the above axiomatization of Ty1 are complete w.r.t. the Henkin semantics of the language, not w.r.t. the standard one.

Thanks to the expressive richness of the language we can also take all the usual logical operators as *defined symbols*; taking only = (for every type) as primitive:

$$\begin{aligned}
&\mathbf{T} \equiv_{Def.} (\lambda y.y =_{(oo)} \lambda y.y)\\
&\mathbf{F} \equiv_{Def.} (\lambda y.y =_{(oo)} \lambda y.\mathbf{T})\\
&\Pi_\alpha \equiv_{Def.} \lambda g.(g =_{(o\alpha)} \lambda a.\mathbf{T})\\
&\wedge \equiv_{Def.} \lambda y.\lambda z.(\lambda h.h(\mathbf{T},\mathbf{T}) =_o \lambda h.h(y,z))\\
&\rightarrow \equiv_{Def.} \lambda y.\lambda z.(y =_o y \wedge z)\\
&\neg \equiv_{Def.} \lambda y.(y =_o \mathbf{F})\\
&\vee \equiv_{Def.} \lambda y.\lambda z.\neg(\neg y \wedge \neg z)
\end{aligned}$$

(where y, z, a, g, h are variables of the respective types $o, o, \alpha, (oo)$ and $((oo)o)$ and we write $h(x,y)$ instead of $(h(x))(y)$ and use the usual infix notation for logical operators).

Andrews [1986] has shown that then we can make do with the following axioms and rule:

$$\begin{aligned}
&H(\mathbf{T}) \wedge H(\mathbf{F}) =_o \Pi_o(\lambda y.H(y))\\
&B =_\beta B' \rightarrow (I(B) =_\alpha I(B'))\\
&(I =_{(\alpha\beta)} I') \rightarrow \Pi_\beta(\lambda x.(I(x) =_\alpha I'(x)))\\
&(\lambda x.B)(A) =_\beta B^{x\leftarrow A}\\
&\iota_\alpha(\lambda x.(x =_\alpha A)) =_\alpha A\\
&A =_\alpha A', B/B^{A\leftarrow A'}
\end{aligned}$$

Here the newly introduced letters H, B', I and I' stand for expressions of the respective types $(oo), \beta, (\alpha\beta), (\alpha\beta)$.

To summarize: we can see the semantics of first-order predicate calculus (which underlies also common model theory) as a matter of connecting expressions to their extensions such that the extension of a whole is always computed from the extensions of parts via a functional application; and we can see Ty1 as the ultimate generalization of this idea.

---

[15] As Andrews [1986] duly points out, this axiom can be seen as a disguised comprehension principle: it involves that to every open formula there corresponds a function, especially that for every formula of type o there is the function which maps all n-tuples satisfying the formula on **T** and all others on **F**.

[16] The ι operator is thus reminiscent both of the 'iota inversum' operator of Russell, and of the ε operator of Hilbert.

[17] For the concept of the Henkin semantics, see, e.g., Andrews [1986] or Shapiro [1991].

3 FAILURES OF EXTENSIONALITY AND CARNAP'S *INTENSION*

As we saw, Carnap pointed out that if what we are after is *meaning* in the intuitive sense, then extensions would not do. We saw that in fact this is rather obvious, for probably nobody would be willing to admit that a truth value is what a sentence *means* and that consequently all true sentences (as well as all the false ones) are synonymous. However, Carnap used more sophisticated counterexamples, mostly concerning predicates. Thus, for instance, he claimed (borrowing from Aristotle's classification) that though the predicates "human" and "featherless biped" are co-extensional (there is, as a matter of fact, no animal species which has no feathers and at the same time has two legs, save us, humans), they are clearly not synonymous. (We all know that a non-human featherless biped is surely conceivable — e.g. a hobbit!).[18]

This means that insofar as a logician wants to explicate the concept of meaning, she must not stay on the level of extensions and must follow Carnap to that of intensions. However, should a logician struggle to explicate meaning at all? Or should she care only about truth, and hence, as Frege maintained, extension? We have seen that Frege endorsed extensions in view of the fact that they are enough to provide for a compositional account for truth; in particular that (i) the assignment of extensions to expressions is compositional; (ii) the assignment of extensions to sentences coincides with the assignment of truth values.

It is important to realize that the Carnapian considerations challenge not only the extensions' capability of explicating the intuitive concept of meaning, but rather also the Fregean way of accounting for truth, by challenging the assumption (i). Consider the sentence

(1) One need not know that all featherless bipeds are human.

It is clearly true: one who has not studied much zoology need not know such thing. However, if extensions are compositional, then we are free to replace the name of an extension by another name of the same extension within the sentence without thereby changing the truth value of the sentence. Hence it seems that also the sentence

(2) One need not know that all humans are human,

which arises out of replacing "featherless biped" by the co-extensional "human", must also be true. But how could anybody sensibly fail to know that humans are humans?

Or consider a sentence of the shape

(3) One plus one is necessarily two.

Such sentences have traditionally been considered as resulting from the application of the operator of necessity to a sentence; hence its extension, i.e. its truth value should be yielded by a combination of the extension of “necessarily” and the extension (truth-value) of “one plus one is two”. It follows that replacing the sentence “one plus one is two” by any other sentence with the same truth value should not change the truth value of (3). However, it is easy to see that if we replace “one plus one is two” by a sentence which is true only contingently, perhaps “Prague is the capital of Czechia”, then the truth value *will* change from **T** to **F**.

---

[18] See [Carnap, 1947, §5].

Hence it is clear that extensions *in general* are not compositional. (What would be Frege’s response to this? Did he not see this? He surely did, but he was convinced we could restrict ourselves to only that part of language where the compositionality of extensions does hold — that this is the ‘logically relevant core’ of language.) Thus a non-extensional logic is needed not only when we are engaged within the philosophical enterprise of explicating the concept of meaning, but also when we want to logically master certain, ‘non-extensional’ contexts.

All in all, it seems that if what we want is either to explicate the concept of meaning, or to account for non-extensional contexts, we should follow Carnap and turn our attention to intensions. However, what is an intension? Is there a possibility of explicating it as explicitly as extension and to establish an ‘intensional logic’?

We saw that Carnap considered two expressions co-extensional if they were equivalent in the precise sense defined in the beginning of the previous section. This yielded him the explication of extensions. Now turning his attention to intensions, he concluded that they can be approached in an analogous way: he proposed to consider two expressions co-intensional if they are *logically* equivalent, i.e. if their equivalence is not simply true, but *logically* true (their equivalence follows by nothing else than the laws of logic). Thus while “the morning star” and “the evening star” are co-extensional, for

(4) The morning star is the evening star

is true, they are not co-intensional, for (4) is not *logically* true. Similarly “featherless biped” and “human” are — for all we know — co-extensional, for

(5) An individual is a featherless biped iff it is human

is true; but they are not co-intensional, for (5) is not *logically* true.

While this meant an advance on the way to explicating co-intensionality, it was not, however, directly on the way to explicating *intension*. But Carnap indicated also a more promising route to the latter goal: he noticed that two sentences are logically equivalent iff they are true w.r.t. the same states-of-affairs. Thus, Carnap [1947], introduced the concept of *state-description*. A state description is a set of atomic sentences containing for every atomic sentence either it, or its negation. Then Carnap assigned to every sentence what he called its *range*: the set of all state-descriptions in which it is true.

Intuitively, state-descriptions represent conceivable states of our world (as Carnap himself puts it, *ibid.*, p. 9, they represent “Leibniz’s possible worlds or Wittgenstein’s possible states of affairs”); formally each of them uniquely determines a maximal consistent set of sentences. Carnap’s observation indicated that the intension of a sentence could perhaps be explicated precisely as its range: as the set of all those state-descriptions (and consequently ‘possible worlds’ — whatever these may be) in which the sentence is true. And precisely this has later become the point of departure of modal and intensional semantics.

4 MODAL LOGIC

In fact logicians did ponder sentences claiming that something is possible or that it is necessary from the very beginning of the enterprise of logic, i.e. since the time of Aristotle. And besides the majority of post-Fregean logicians, who developed the wonderfully simple and transparent extensional logic outlined by Frege (and also Boole and others), there was always a minority of those who thought about ways of integrating even the non-extensional contexts into logic.

In 1932 Lewis and Langford published their *Survey of Symbolic Logic*, in the Appendix of which C. I. Lewis presented five axiomatic systems of modal logic (thereafter known as S1–S5). The crucial logical connective which appeared in them was the so-called *strict implication* ($\prec$); however, the modern modal logicians came to the conclusion that as this is one of the triad of mutually interdefinable modal operators, the other members of which being *necessity* ($\Box$) and *possibility* ($\Diamond$), we are free to take any one of the three as the primitive one and mostly settled for $\Box$. ($\Diamond S$ then can be defined as $\neg\Box\neg S$, whereas $S_1 \prec S_2$ as $\Box(S_1 \rightarrow S_2)$). Therefore most modern axiomatic systems of modal logic are based on the necessity operator (usually called the *box*; the possibility operator being called the *diamond*).[19] It is, however, good to realize that there is a respect in which Lewis' strict implication is more important than its two relatives: while the necessity and the possibility operators do not correspond to anything terribly important in ordinary discourse,[20] the strict implication seems to correspond to the all-important contrafactual conditional of natural language.

It was clear that the language of modal logic did not allow for an extensional interpretation. $\Box$ was syntactically on a par with $\neg$, hence within the extensional framework it would have to denote a unary truth function. Assume that such a function exists; call it $f_\Box$. And assume that within the language we are interested in there is a sentence, say "1=1", which is necessarily true, and another sentence, perhaps "Prague is the capital of Czechia", which is true, but not necessarily (if there were no such two sentences, the situation would be trivial). Then it must be the case that

$$\begin{aligned}&\mathbf{T} = ||\Box(1 = 1)|| = f_\Box(||1 = 1||) = f_\Box(\mathbf{T}) =\\ &= f_\Box(||\text{Prague is the capital of Czechia}||) = ||\Box(\text{Prague is the capital of Czechia})|| =\\ &= \mathbf{F}.\end{aligned}$$

Hence, in pain of contradiction, no such function exists.

This means that if modal logic is to be interpreted, we need a semantics which is not extensional. The first attempts to build a feasible semantics were presented in the fifties;

---

[19] For a detailed exposition of modal calculi, see Hughes & Cresswell [1968], Chellas [1980] Gamut [1991], or Blackburn *et al.* [2000].

[20] This was argued for by Quine. He pointed out [1992, 73] that when, in ordinary discourse, we apply the adverb "necessarily" to a sentence, we usually do not mean to express anything of the kind studied by modal logic, but rather either want to express that the sentence is "presumed acceptable to our interlocutor and stated only as a step toward the consideration of moot ones" or want to "identify something that follows from generalities already expounded, as over against new conjectures or hypotheses". Thus, Quine concludes that "expression [of necessity] serves a purpose in daily discourse, but a shallow one."

they culminated in the work of Saul Kripke, who is nowadays usually considered as *the* author of such a semantics.[21] Let us consider his proposals in detail.

Kripke came to the conclusion that we must let sentences denote not truth values, but rather subsets of a given set. He called elements of the underlying set *possible worlds*, which made his proposal (treacherously) easy to grasp: each sentence is taken to denote the set of those possible worlds in which it is true. This further lets us explicate necessity as 'truth in every possible world' and possibility as 'truth in at least one possible world', which is again very plausible. Formally, given a set $W$of possible worlds, we can define $f_\Box$ as follows:

$$\begin{aligned} f_\Box(x) &= W \text{ iff } x = W \\ &= \emptyset \text{ otherwise.} \end{aligned}$$

This semantics yields a rather plausible class of tautologies, like $\Box S \rightarrow S$ ('if something is necessarily the case, then it is the case'), $\Box S \rightarrow \Box\Box S$ ('if something is necessarily the case, then it is necessary that it is necessarily the case') etc.

Aside of the story about 'possible worlds', there is also a purely algebraic way to understand Kripke's proposal. The ordinary (extensional) logical connectives make the set of denotations of sentences into a Boolean algebra (with conjunction acting as the meet, disjunction as the join and negation as the complement). As there is no need for a more complicated Boolean algebra than the simplest, two-element one, we make do with the two truth values. However, the introduction of the modal operators blocks this simplest possibility; so the solution is to settle for less trivial Boolean algebras, and as every Boolean algebra can be represented by the algebra of subsets of a set, we have the semantics outlined above.[22]

In fact, Kripke proposed a more sophisticated semantics,[23] in which the set of possible worlds was supplemented by a binary relation, which is called the *accessibility relation*. The idea is that what would be relevant for the necessary truth of S w.r.t. a world w is not the truth of S w.r.t. all worlds whatsoever, but rather only its truth w.r.t. those which are *accessible* from w. Thus, if we are, for example, trying to explicate *physical* necessity, then we might, considering necessity in a world w, want to disregard those worlds in which there are different physical laws (and hence we make the accessibility relation into the relation of *sharing the same physical laws*).

It has turned out that by fine-tuning the accessibility relation we can develop sound and complete semantics for a rich variety of axiomatically defined modal logics. In particular, it has turned out that many axioms of modal logic correspond to simple properties of the accessibility relation. Thus, for example, $\Box S \rightarrow S$ holds if and only if the relation is reflexive; and similarly for many other axiom-candidates. Investigations into this kind of correspondence has generated what is now called the *correspondence theory*.

There is a further way of understanding the step from the extensional to the intensional, possible-world semantics. We can imagine that we simply change the extensional semantics so as to admit that the 'actual' distribution of truth values among sentences is not the only one possible — hence that we have a whole set of acceptable truth valuations.

---

[21] See Copeland (2002) for a detailed overview of the emergence of possible-worlds-semantics.

[22] Cf. Peregrin (2005a).

[23] See [Kripke, 1963; 1965].

Each such valuation corresponds, on the one hand, to a Kripke possible world, while, on the other, can be seen as amounting to a Carnap *state-description*, as a means of pointing out certain sentences. In this way, Kripke's proposal come to connect with the one of Carnap.[24]

5 MODAL PREDICATE LOGIC

There is a limited parallel between modal propositional logic and extensional predicate logic. We may compare possible worlds with individuals and modalities with quantifiers: indeed in the simplest case,

> necessarily S

means

> for (in) every world, S

whereas

> possibly S

means

> for (in) at least one world, S.

In the general case the situation is more complicated, but still we can say that

> necessarily S

means

> for (in) every accessible world, S

whereas

> possibly S

means

> for (in) at least one accessible world, S.

This opens the way for a reduction of some modal logics to the extensional first-order predicate logic.[25] (For other systems of modal logic this may not be possible because some modal formulas correspond to such properties of the accessibility relation that are not first-order definable.[26])

---

[24] An approach to the semantics of modal logic alternative to that of Kripke and more congenial to that of Carnap was fostered by [Hintikka, 1969]. Its basic concept of *model set* is quite close to Carnap's *state description*.

[25] As we have already pointed out, $\Box S \to S$, for instance, is valid within a modal logic iff the underlying accessibility relation is reflexive, i.e. iff, denoting the relation as BR, $\forall w.wBRw$.

However, once we have a propositional logic of possibility and necessity, we would want to extend it to a predicate logic. On first sight, this should be simply a matter of adding the necessity operator to the standard predicate calculus; but in fact the situation is somewhat trickier. Whereas in extensional predicate calculus we have the universe of individuals and in modal propositional calculi we have the universe of possible worlds, which can be seen as, in a sense, replacing it, in modal predicate logic we have *both*, and we face the problem of their mutual relationship.

This is indeed a problem, though *prima facie* the situation might seem to be clear: it appears that the universe of individuals must be relativized to possible worlds, that there should be a separate universe for each world. However, the situation is not so simple. First, this would imply considerable complications for the formal apparatus. The point is that whereas the formula $\Box\forall xS$ ("for every world and every its individual, S") would continue to make sense, $\forall x\Box S$ would become senseless, for it would require us to quantify over individuals which do not belong to any possible world, which would make no sense.

Besides this, there is a less formal problem. One of the basic points of modal logic is the analysis of contrafactual locutions, which in natural language typically have the form "If [it were the case that] ..., then [it would be the case that] ...". This means that one of the prototypical kinds of sentences which we might hope to capture by means of modal predicate calculus is that exemplified by

(6) If I were the president of Russia, I would make St. Petersburg the capital.

And it seems that this requires that *the same* individual be in *more than one* possible world: namely that there is a possible world in which I, the same person which exists in the actual world, am the president of Russia.

There are a number of responses to this problem. Thus, D. Lewis [1968; 1986], for example, wanted to save the idea of the world-specific universes by postulating a *counterpart relation* among individuals of the universes of different worlds. Hence though I do exist only in the actual world, in other worlds there are my counterparts which are, for the purposes of semantic analysis, indistinguishable from me.

Tichý [1971], on the other hand, accepted that individuals are prior to possible worlds: which made him reduce individuals to bare 'property-hangers'. That is, the individuals outside of possible worlds do not have any nontrivial properties (though they are 'numerically' distinct from each other), and they acquire such properties only within the context of possible worlds: and possible worlds then emerge from a distribution of some basic properties among the transcendental individuals.

However, also this solution has consequences which appear unwanted. It implies, for example, that there cannot be a possible world in which I do not exist — in some worlds I may be a stone or an ash-tray (or perhaps even a summer breeze?), but there is for me no way of not being there at all. It also implies that there are worlds which do not differ in any humanly recognizable way, but which *are* different in that, for example, the property-hanger underlying a person A in one of them underlies a person B in the other and *vice versa*.

---

[26] Thus, for example, to $\Box\Diamond S \rightarrow \Diamond\Box S$ there arguably corresponds no property of the accessibility relation which would be expressible within the language of FOPC (see Goldblatt, [1975]). Another example is $\Box(\Box S \rightarrow S) \rightarrow \Box S$ (see [Boolos, 1979, 82].

These two extreme approaches may be partly reconciled by means of differentiating two senses of "exists": there is a broader sense, in which every individual which can be found in any possible world exists (*simpliciter*); and there is a narrower sense in which existence is relativized to a possible world and in which only what can be found in a world exists in that world. Some individuals which exist in the broader sense may not exist in the narrower sense in a given possible world. Existence-in-a-world then can be conceived of as a *property*, a property which in every possible world is instantiated by those individuals which exist-in-the-world.

The way we choose to construe the relationship between the universe of individuals and that of possible worlds also bears on the validity of the so-called Barcan formula. This formula codifies the interchangeability of the quantifications over the two universes:

$$\forall x \Box F \leftrightarrow \Box \forall x F.$$

If the universes are treated as independent (in the sense that we take all individuals to exist in every possible world — so that the quantification over individuals is always taken to be over the whole universe), the formula is unproblematically true. On the other hand, if the individuals are relativized to possible worlds (so that the quantification over individuals within a possible world is only over the individuals which exist in that possible world, which need not be all individuals), then this formula, if not deemed utterly meaningless (see above), may hold only for such interpretations in which the universe of individuals of every possible world would be the same as that of every possible world accessible from it.

It seems that from the intuitive viewpoint, it would be natural to have both the possibility of the same individual occurring in more than one world and the possibility of an individual present in a world being utterly absent from another world. However, this would obviously lead to an apparatus much more complicated and much less elegant than the above two.[27]

6 MONTAGUE'S 'LOCALLY' INTENSIONAL LOGIC

The most famous system of modal predicate logic was presented by Richard [Montague, 1970a; 1970b; 1973]. He called the system *intensional logic* and this term has been almost universally accepted, so that the usage of the term "modal" is now normally restricted to propositional logic. The basic idea was that of *explicating intensions generally as functions from possible worlds to extensions*. This was compatible with how we could see the relationship between the extension of a sentence and its Kripkean intension: the intension of a sentence could be seen as a function mapping each world on the truth value of the sentence for that world.

---

[27] See [Gamut, 1991, §3.3] for a further discussion of the choice between an absolute universe and relative universes. For discussions of further aspects of the 'metaphysics of possible worlds' see Loux [1979] or Divers (2002).

This means that while an intension of a sentence is a function from possible worlds to truth values, that of an individual term is a function from possible worlds to individuals and that of a predicate is a function from possible worlds to classes of individuals or classes of n-tuples of individuals. Thus, the intension of "the president of the USA" is a function mapping every possible world on its president of the USA (if any); whereas the intension of "featherless biped" is a function mapping every world on the class of its featherless bipeds. In this way, Carnap's idea of ranges gets generalized to expressions of all categories.

However, this elegant solution brings with it a grave problem. We saw that within extensional semantics the uniform way of combining denotations of parts into the denotation of a whole was functional application; but if we explicate intensions as Montague did, this is no longer possible. While we can apply the extension of "featherless biped" to the extension of "the president of the USA" to yield us the extension of "The president of the USA is a featherless biped" (for the former extension is a function from individuals to truth values and the latter is an individual), we cannot do the same with the respective intensions — the intension of "featherless biped" is no longer the kind of function which would be applicable to the intension of "the president of the USA".

What we can do is to take the values (i.e. extensions) of these intensions for a particular world and let them yield us the extension of "The president of the USA is a featherless biped" for the possible world. And we can do this for any possible world. Hence as the intension of the sentence is uniquely determined by its truth values w.r.t. all possible worlds, the intension can be obtained by obtaining the extensions for every possible worlds. This was Montague's strategy: he gave the rules for computing extensions assuming that their totality yields us also intensions.

Obviously, the situation is not so simple that we could always merely take extensions of components and use them to yield us the extension of the corresponding compound — if so, then the difference between extensional and intensional logic would be rather trivial. The *raison d'être* of intensional logic, we saw, was the fact that in some cases we may need *more* than just the extension of a component to get the extension (not to speak about intension) of the compound.

Montague's solution was that (i) each expression of his logic had both an extension and an intension; (ii) in extensional contexts expressions continued to denote their extensions; and (iii) there was a mechanism which would take care of intensional contexts in that it would allow, in effect, an expression to 'exceptionally' denote its intension instead of the extension. Thus, Montague's was what we could call a *locally* intensional logic — intension enters the scene only where there is no way of making do with extension. The mechanism with which Montague accomplished this was realized by an operator, namely $^\wedge$. Its role was such that if we denote the extension of an expression e as $||e||_E$ whereas its intension as $||e||_I$, we can write

$$||^\wedge e||_E = ||e||_I.$$

Thus, as "necessarily", as we saw, constitutes an intensional (i.e. non-extensional) context, the formula corresponding to the natural language *necessarily S* will not be $\Box S$, but rather $\Box{}^{\wedge}S$. And while the extension of S is a truth value, that of ${}^{\wedge}S$ equals the intension of S and hence is a set of possible worlds.

Let us describe Montague's system in greater detail.[28] However, we will present it in a slightly simplified form and also using a different notation than Montague, in order to stress the continuity with Church's Ty1 described above.[29]

Montague's intensional logic (MIL)

Syntax

ι and o are basic types, and if α and β are types then also $(\alpha\beta)$ is a type. Moreover, if α is a type, then also $(\alpha\omega)$ is a type. For each type we have an unlimited stock of constants and variables. If a is an expression of type $(\alpha\beta)$ and b an expression of type β, then $a(b)$ is an expression of the type α; and if a is an expression of type α and x a variable of type β, then $\lambda x.a$ is an expression of type $(\alpha\beta)$. Moreover, if a is an expression of type α, then ${}^{\wedge}a$ is an expression of type $(\alpha\omega)$; and if a is an expression of type $(\alpha\omega)$, then ${}^{\vee}a$ is an expression of type α.

Semantics

To each type α there correspond two domains S_α and D_α, where S_α is identical to $D_{(\alpha\omega)}$. D_o is the set of the two truth values, D_ι is a universe of discourse (any given set); $D_{(\alpha\beta)}$ is a set of functions from D_β to D_α, and $D_{(\alpha\omega)}$ is a set of functions from a set of 'possible worlds' (any given set) to D_α. An interpretation maps every constant a of type α on an element of S_α; a valuation maps every variable x of type α on an element of D_α. If I is an interpretation and V a valuation, then for every expression a of the language and every possible world w we define the extension $||a||^w_{I,V}$ of a in w, thereby defining the intension $||a||_{I,V}$ of a, as the function mapping every w on $||a||^w_{I,V}$. For a constant c, $||c||^w_{I,V} = (I(c))(w)$; for a variable v, $||v||^w_{I,V} = V(v)$. $||a(b)||^w_{I,V} = ||a||^w_{I,V}(||b||^w_{I,V})$ and $||\lambda x.a||^w_{I,V}$ is the function which maps every element i of D_β on $||a||^w_{I,V'}$, where V' is just like V with the single possible exception that it maps x on i. Moreover, $||{}^{\wedge}a||^w_{I,V}$ is the function which maps every possible world w' on $||a||^{w'}_{I,V}$; whereas $||{}^{\vee}a||^w_{I,V}$ is $||a||^w_{I,V}(w)$.

As an example, consider the sentence

(7) Madonna sings

In terms of Ty1 (or, for that matter, FOPC), it might be analyzed simply as

(7′) **sing**(**Madonna**)

where **sing** is an expression of the type $(o\iota)$ and **Madonna** of the type ι. Now passing over to MIL, nothing needs to change; save for the fact that each of the expressions as well as the whole sentence will now have also also an intension; the intension of the sentence being denoted by $^\wedge(\mathbf{sing}(\mathbf{Madonna}))$.

---

[28] See [Mongague, 1974]; and also [Gallin, 1975], [Partee, 1976] and [Gamut, 1991, Chapter 6].

[29] Montague used the letters e, t and s instead of ι, o and ω, and he also wrote $\langle b, a\rangle$ instead of $(\alpha\beta)$. Hence, for example, his equivalent of $(o(o\iota))$ was $\langle\langle e, t\rangle, t\rangle$.

To illustrate some idiosyncrasies of Montague's approach, consider a slightly more sophisticated example:

(8) John finds a unicorn

Within FOPC, the most straightforward analysis would be

(8′) $\exists x\,(\mathbf{find}(\mathbf{John},x) \;\wedge\; \mathbf{unicorn}(x))$.

Switching to Ty1, we need to make the binary predicate **find** into an expression of the type $((o\iota)\iota)$ (taking a term to a unary predicate):

(8^2) $\exists x\,(((\mathbf{find}(x))(\mathbf{John})) \;\wedge\; \mathbf{unicorn}(x))$.

Using the mechanism of lambda-abstraction, this can be further transformed to

(8^3) $(\lambda y.\exists x(((\mathbf{find}(x))(y)) \wedge \mathbf{unicorn}(x)))(\mathbf{John})$.

Now if we define (where r is a variable of the type $(o(o\iota))$)

$$\mathbf{find^*} \equiv_{Def.} \lambda r.\lambda y.r(\lambda x.(\mathbf{find}(x))(y)),$$

this can be further turned into

(8^4) $(\mathbf{find^*}((\lambda p.\exists x(p(x) \wedge \mathbf{unicorn}(x)))))(\mathbf{John})$

Here **John** is a constant of the type ι, x is a variable of the same type, p is a variable of the type $(o\iota)$, and **unicorn** and **find*** are constants of the types $(o\iota)$ and $((o\iota)(o(o\iota)))$, respectively. (Note that the step from **find** to **find*** was a purely technical one, relying on the fact that an individual, or any other object, can be identified with the class of all classes to which it belongs; hence though the type of the counterpart of "find" should be intuitively $((o\iota)\iota)$, **find*** is of the type $((o\iota)(o(o\iota)))$.)

Now consider a similar sentence

(9) John seeks a unicorn,

which, however, differs from (8) in the crucial respect that the object position of "seek", unlike that of "find", constitutes an intensional context. (Whereas you cannot find a unicorn without there being one, you can perfectly well seek a unicorn even when no unicorns exist.) Thus the counterpart of "seek", unlike that of "find", should be intuitively of the type $((o\iota)(\iota\omega))$, but due to the maneuver explained above we have **seek*** of the type $((o\iota)(o(o(\iota\omega))))$. Hence we must use, instead of p, a variable q of the type $o(\iota\omega)$ and, consequently, 'intensionalize' its argument:

(9′) $(\mathbf{seek^*}((\lambda q.\exists x((q)(^\wedge x) \wedge \mathbf{unicorn}(x)))))(\mathbf{John})$.

Montague's own analysis only differs from this one in two minor points: first, he lets *all* arguments of *all* predicative expressions undergo intensionalization (which forces different types of the predicative expressions); and, second, he uses variables which have only extensions and no intensions. This results into the formula[30]

$$(9^2)\ \ (\mathbf{seek^*}(\lambda q.\exists u(\mathbf{unicorn}({}^{\wedge}x) \wedge ({}^{\vee}q)({}^{\wedge}x))))({}^{\wedge}\mathbf{John}),$$

where q is a variable of the type $((o(\iota\omega))\omega)$, and **unicorn** and **seek*** are constants of the types $(o(\iota\omega))$ and $((o(\iota\omega))(o((o(\iota\omega))\omega)))$, respectively. It is clear that only to comprehend the type of the verb is anything but easy, and to decipher the whole formula is even harder.

An axiomatization of Montague's intensional logic was presented by [Gallin, 1975]. It turns out that there are only two axioms which must be added to an axiomatization of Ty1, namely

$$\Box({}^{\vee}x = {}^{\vee}y) \rightarrow (x = y);$$
$${}^{\vee\wedge}x = x,$$

where

$$\Box x \equiv_{Def.} ({}^{\wedge}x = {}^{\wedge}\mathbf{T}).$$

7 'GLOBALLY' INTENSIONAL LOGIC

Thus Montague still declares extensions as the basic semantic values, and takes intensions to be relevant only in that they can temporarily assume the place of extensions. Therefore he needs the operator $^{\wedge}$, which is, however, not unproblematical in that it is, in contrast to usual logical operators, *not* defined compositionally. In particular, there is no function which would take us from the semantic value (extension) of e to that of ${}^{\wedge}e$. This is understandable in that insofar as Montague sees extensions as the basic semantic values, intensionality cannot but be a failure of compositionality. Note however that though $^{\wedge}$ is generally applicable, there is no way of defining it generally: going from $||e||$ to $||{}^{\wedge}e||$ is not a matter of a rule or of an algorithm, but rather of the advance knowledge of $||{}^{\wedge}e||$.

We can say that facing the dilemma of either taking intensions at face value (i.e. regarding them directly as denotations of sentences) or saving the Fregean paradigm (according to which one constituent of every complex expression denotes a function applicable to the denotations of the others), Montague voted for the second alternative and hence treated expressions as denoting extensions, letting intensions do their job only via the mediation of the operator $^{\wedge}$. However, it seems that to embrace the first alternative may be in some respects better. We may let expressions denote directly intensions (thereby achieving what can be called *globally* intensional logic) and modify the Fregean paradigm.

Let f be a function from D_ω to $D_{(\alpha\beta)}$ and let g be that from D_ω to D_β. Let us use the term *intensional application* of f to g for the operation which produces the function h from D_ω to D_α such that for every $x \in D_\omega, h(x) = (f(x))(g(x))$. It is easy to see that if we change Montague's semantics in such a way that every expression comes to denote its intension and every application of a denotation to other denotations becomes an *intensional* application, everything will work as smoothly as before.

---

[30] See [Montague, 1973].

Moreover, it takes only a minor enhancement of the language of MIL to be able to articulate intensional application explicitly. If we add variables ranging over possible worlds, then obviously the result of the intensional application of the denotation of an expression A of the type $((\alpha\beta)\omega)$ to an expression B of the type $(\beta\omega)$ will be denoted by $\lambda w.A(w)(B(w))$. Thus, the intension of (9) will be denoted by (where for the sake of comprehensibility we write the argument w as a subscript):

$$(9^3)\ \lambda w.(\mathbf{seek*}_w(\lambda w.\lambda q.\exists x(\mathbf{unicorn}_w(x) \wedge q(x))))(\mathbf{John})$$

where the respective types of q and **unicorn** have been changed to $((o\iota)\omega)$ and $(o\iota)$, whereas that of **seek*** to $(((o\iota)((o(o\iota)\omega)))\omega)$.

In what respect may a globally intensional logic be considered as preferable to the locally intensional one? First, there is some slight simplification both of the types corresponding to the common expressions of natural language and of the formulas which emerge from their analyses. Second, we get rid of Montague's problematic operators $^\wedge$ and $^\vee$. And third, we can take the theoretical relation of denotation as a reasonable explication of the pre-theoretical relation of meaning, for now it relates expressions to their *intensions*, which, unlike extensions, are capable of serving as explicata of meanings (though not as perfect ones).[31]

A simple way to achieve globally intensional logic is to elevate possible worlds to a fully-fledged type. Hence we have, in addition to the basic types o and ι, also the type ω of possible worlds; and of course all the types which can be obtained from all the three. And we assume that each expression is mapped simply on its intension (a function from possible worlds to the corresponding extensions).

This logic, within the context of post-Montagovian research, was presented by [Gallin, 1975] under the name of *two-sorted type theory* (or Ty2, as a generalization of Church's system with a second domain, besides that of individuals).[32] However, it should not escape our attention that an intensional logic based on this approach was presented independently by [Tichý, 1975; 1978a; 1978b].

Ty2 or Tichý's intensional logic

Syntax

ι, o and ω are basic types, and if α and β are types then also $(\alpha\beta)$ is a type. For each type we have an unlimited stock of constants and variables. If a is an expression of type $(\alpha\beta)$ and b an expression of type β, then $a(b)$ is an expression of type α; and if a is an expression of type α and x a variable of type β, then $\lambda x.a$ is an expression of type $(\alpha\beta)$.

---

[31] See the last section.

[32] For a thorough discussion of the relationship between MIL and Ty2, see [Zimmerman, 1989].

Semantics

To each type α there corresponds the domain D_α. D_o is the set of the two truth values, D_ι is a given universe of discourse and D_ω is a given set of possible worlds;[33] $D_{(\alpha\beta)}$ is a set of functions from D_β to D_α. An interpretation maps every constant of type α on an element of D_α; a valuation maps every variable of type α on an element of D_α. If I is an interpretation and V a valuation, then for every expression a of the language we define the intension $||a||_{I,V}$ in the following way. For a constant c, $||c||_{I,V} = I(c)$; for a variable v, $||v||_{I,V} = V(v)$. $||a(b)||_{I,V} = ||a||_{I,V}(||b||_{I,V})$ and $||\lambda x.a||_{I,V}$ is the function which maps every element i of D_β on $||a||_{I,V'}$, where V' is just like V with the single possible exception that it maps x on i.

We can see that Ty2 is merely a minor variation on Ty1. However, we saw that Ty1 was purely extensional; and now it seems that intensional logic boils down to its simple variation — a variation which merely adds one more type and does not affect its basic extensionality in any substantial way. Should we take this as a *reductio ad absurdum* of the possibility of a ('truly') intensional logic? Not really; but we should take it as an indication that a deeper insight into the concept of *intensionality* is needed.

At first, we must distinguish dealing with an abstract, mathematical structure from dealing with something via the structure.[34] It is true that Ty2, in itself, is in no clear way any more intensional than Ty1. However, note that before we can consider a formal system as a *language* and especially as a language underlying a *logic*, we have to single out the category of sentences, and more generally the categories which are to regiment the pre-theoretic categories of expressions of our ordinary language. And here is where an important difference emerges: whereas for Church sentences were expressions of the category o, for an intensional logician they are rather expressions of the category $(o\omega)$. And this difference spreads to the other categories.

Hence the step from an extensional language to an intensional one does not consist in changing the structure of Ty2, but rather in the way of employing it for logical purposes, in the way of matching it with the natural language and pre-formal reasoning.

8 QUINE'S EXTENSIONALIST PROGRAM

Rampant philosophical objections to intensions and intensional logic were raised by W. V. Quine [1960]. The basic objection was that intensions were not sufficiently clearcut to be included into the subject matter of a science so rigorous as logic. According to Quine, they do not have real boundaries: there is no clear telling where one ends and another begins. Take propositions, the intensions of sentences: is the proposition expressed by "Berlin is east from Paris" the same as that expressed by "Paris is west of Berlin"? Who is to decide? Similarly, is the intension of the predicate "to be bigger than John" the same as "not to be smaller than John"?

---

[33] Tichý takes the language of his logic to be fully interpreted, so he does not count with varying universes. (He sees himself as the continuator of the Frege–Russell tradition as discussed by [Goldfarb, 1979].

[34] See Peregrin (2000b).

Quine concluded that intensions are simply illusory: there cannot be an entity, he insists, which would not have a clear-cut boundary. (His famous slogan was "No entity without identity.") Hence propositions, properties and intensions in general are non-entities. In comparison, individuals, truth values and sets of individuals, which underlie extensional logic, are objects *par excellence*. At least, that is what Quine claims.

However, did not Carnap and Kripke help us to a *rigorous* concept of intension, which *is* as clearcut as anything can be clearcut at all? True, meanings of expressions of natural languages are fuzzy, but Kripkean intensions are *functions*, and functions are in general well-defined if anything is. It may be an obscure matter whether two sentences of natural language denote the same intension; but intensions themselves are not obscure! Hence are Quine's scruples simply preposterous?

Not really. In fact, the fuzziness Quine diagnosed did find its way even into our notion of intension, hidden within the Trojan horse of the concept of possible world. Is there a possible world in which Berlin is east from Paris, while Paris is *not* west of Berlin? If we imagine possible worlds in some intuitive (in the Kantian sense) way, then it seems not, but could we not have possible worlds with some bizarre non-Euclidean geometries in which something like this would be possible? (And should we insist on *intuitivity* of possible worlds at all?) And again, who is to decide?

What if we construe possible worlds as maximal consistent classes of sentences? Are such worlds not automatically excluded at least in this case? Are not "Berlin is east from Paris" and "Paris is not west of Berlin" incompatible? Well, if all expressions already have exactly specified meanings, then it is determined what is incompatible with what and hence what are the maximal consistent classes of sentences. However, in such a case it is unclear what could be the reason for taking pains to build a possible-worlds semantics: this is an achievement for a language whose semantics is in need of explication, i.e. for a formal language determined merely in terms of axioms and inference rules or for a natural language existing only via the practices of certain communities.

Besides this, even if we denied that the Carnapo-Kripkean intensions are fuzzy, a problem would persist. What we are interested in are *the very intensions of our expressions*; and to promote the goal of their explication it is of little help to have some crisp intensions, about which we are nevertheless unable to say by which real expressions they are denoted. (Lewis, [1975], pointed out that it is one thing to study abstract languages, and it is another thing to pin down the abstract language which can be identified with an empirical language. Our point here is that if what we are interested in is meaning within the empirical languages, then the former enterprise in itself is not of much use.)

Hence Quine seems to have a good point after all: intensions *are* essentially fuzzy. However, be they as fuzzy as they may, the two points stated above which render extensional semantics problematic persist: it is incapable of providing (i) an adequate explication of meanings and (ii) an adequate analysis of non-extensional contexts. What is Quine's response to these obstacles?

As for (i), Quine simply rejects the concept of meaning. This is not to say that he rejects that our expressions are *meaningful*, he only insists that such meaningfulness is a more-or-less matter and as such is not reasonably graspable as a possession, by the expression, of a definite object. "I would not seek," claims Quine [1992, 56], "a scientific rehabilitation of something like the old notion of separate and distinct meanings; the notion is better seen as a stumbling block cleared away."

An objection which can be raised against this view is that the fact that something is fuzzy does not automatically imply that it could not be useful to explicate (or model) it in a non-fuzzy way. After all, everything we encounter within the empirical world is (more or less) fuzzy, and we often profit from making idealized models.[35] The whole point is to be aware of the level of idealization employed and not to employ the model where a higher resolution is required.

As for (ii), Quine claims that some of the alleged non-extensional contexts may be disregarded for their role is in fact marginal; whereas others can be dealt with in some indirect extensional ways. Thus, the contexts studied by ordinary modal logic are, according to him, of the first kind: the contexts *possibly* ... and *necessarily* ..., as understood by modal logicians, are not really important for natural language.[36]

But there are less marginal contexts, like those of propositional attitude reports (without which, for example, no psychological theory would be imaginable). How does Quine want to cope with them? What he proposes is to understand such reports as expressing a relation between an individual and a *sentence*. Thus "John believes that all featherless bipeds are human" will get analyzed as a relation between John and the *sentence* "all featherless bipeds are human". (What if John does not speak English? Would it still make sense to say that his believing that all featherless bipeds are human is a matter of his relationship with an *English* sentence? In one sense yes: he would be related to the sentence by the relation which relates a person to a sentence iff the latter expresses the belief of the former.)

It is now up to the reader to weigh the fuzziness of intensions against the cumbersomeness of the Quinean analysis of intensional contexts.

9 BEYOND INTENSIONS

Anyway, intensional logic has proved itself especially fruitful for the purposes of explicating meanings in natural language[37] (simultaneously with the systems of Montague and Tichý there also appeared others, like that of [Creswell, 1973]). However, already from the outset it became clear that even intensions are not entirely sufficient for this purpose. This was particularly apparent in the case of propositional attitude reports (mentioned already above). Consider

(10) John believes that one plus one equals two

Viewed from the perspective of intensional semantics, this statement claims that there is a relation of *believing* between a person John and a proposition that one plus one equals two. And the proposition can be nothing else than an intension, namely the intension of "one plus one equals two", hence the class of all possible worlds. So (10) claims that John is in the relation of believing to the class of all possible worlds. But exactly the same is claimed by any sentence which results from (10) by replacing the object clause by any other mathematical truth, i.e.

---

[35] See Peregrin (*ibid.*).

[36] See footnote 20.

[37] See [Partee and Hendriks, 1997].

(11) John believes that every consistent first-order theory has a denumerable model.

This would entail that (10) and (11) are synonymous and especially that they cannot differ in truth value. But while it is unlikely that someone does not believe that one plus one equals two, it is surely possible that he has no idea about properties of first-order theories.[38]

There are several responses to this observation. One of them is not to tie possible worlds to *empirical* possibility. Thus Hintikka [1978] proposed what he called 'impossible possible worlds': such a world is not possible in the sense of being realizable, but is possible in the sense that somebody might simultaneously believe everything which holds in it.

Another, more popular response was to supplement intensional semantics by some superstructure which would allow us to explicate meanings as something more 'fine-grained' than intensions (then we can speak, together with Cresswell [1975], about *hyperintensional* semantics and perhaps logic). Thus David Lewis [1972], following a hint of Carnap [1947], proposed seeing meanings of compounds as a kind of structures based on the syntactic structures of the corresponding expressions, but involving intensions of their components. These ideas were then elaborated especially by Cresswell [1985]. Tichý [1986], in a similar vein, proposed seeing meanings of compounds as 'constructions' of their intensions from the intensions of their parts: hence to see the meaning of *one plus one* as the construction of the number two out of the operation of addition and two instances of the number one; and to see the meaning of *one plus one equals two* as the construction of the truth value **T** out of this construction, the relation of equality and another (trivial) construction of the number two. And propositional attitudes are then construed as relations to constructions.

This is connected to another problem concerning the Carnapian approach to intensions. Having followed Carnap's explication, whereby intensions became in fact extensions relativized to the state of the world, we have arrived at an extension–intension distinction which makes a nontrivial sense only in the case of empirical terms. Where there is no dependence on the state of the world, there the distinction between extension and intension is trivial (the intension becomes a *constant* function mapping every possible world on the same extension). However, it seems that the intuition underlying the Carnapian distinction applies also to mathematical discourse. Hence, should not the distinction between intension and extension apply also to non-empirical vocabulary?

The point behind the 'intuitive' extension and intension (which Carnap set out to explicate) seems to be that whereas on the extensional level we can only say what (actually) 'falls under' a word, on the level of intension we can say *why* it is that it falls under it. The Carnapian explication exploits the idea that the *knowing why* can be explicated as *knowing the way extension depends on the state of the world.* We can know what the sum of two numbers is without knowing why; and the knowing why, understanding what addition is, can possibly be equated with the ability to give the sum of *arbitrary* numbers. Similarly, knowing why some things fall under a term, knowing the corresponding concept, is equated with the ability to give the extension of the term w.r.t. arbitrary circumstances.

---

[38] The problems posed by the propositional attitude reports were pointed out already by Carnap [1947]; in the post-Montagovian era the discussion was revived by [Lewis, 1972; Partee, 1982; Cresswell, 1985] and others.

From this viewpoint, it would seem that minimally within the realm of mathematics, it is something of the kind of Tichyan constructions, rather than Carnapian intensions that should play the role of meanings in the intuitive sense.[39]

ACKNOWLEDGEMENT

Work on this text was supported by the reseearch grant No. 401/04/0117 of the Grant Agency of the Czech Republic.

BIBLIOGRAPHY

[Ajdukiewicz, 1935] K. Ajdukiewicz. Die syntaktische Konexität, *Studia Philosophica* 1, 1-27, 1935.

[Andrews, 1986] P. B. Andrews. *Introduction to Mathematical Logic and Type Theory*, Academic Press, San Diego, 1986; second edition Kluwer, Dordrecht, 2002.

[Bar-Hillel, 1950] Y. Bar-Hillel. On syntactical categories, *Journal of Symbolic Logic* 15, 1-16, 1950.

[Bar-Hillel, 1953] Y. Bar-Hillel. A quasi-arithmetical notation for syntactic description, *Language* 29, 47-58, 1953.

[van Benthem and ter Meulen, 1997] J. Benthem van and A. ter Meulen. *Handbook of Logic and Language*, Elsevier / MIT Press, Oxford / Cambridge (Mass.), 1997.

[Blackburn *et al.*, 2000] P. M. Blackburn, M. de Rijke and Y. Venema. *Modal Logic*, Cambridge University Press, Cambridge, 2000.

[Boolos, 1979] G. Boolos. *The Unprovability of Consistency*, Cambridge University Press, Cambridge, 1979.

[Carnap, 1947] R. Carnap. *Meaning and Necessity*, University of Chicago Press, Chicago, 1947.

[Casadio, 1988] C. Casadio. Semantic categories and the development of categorial grammars, in *Categorial Grammars and Natural Language Structures* (ed. T. Oeherle et al.), Reidel, Dordrecht, 95-123, 1988.

[Chellas, 1980] B. F. Chellas. *Modal Logic: An Introduction*, Cambridge University Press, Cambridge, 1980.

[Church, 1940] A. Church. A formulation of the simple theory of types, *Journal of Symbolic Logic* 5, 56-68, 1940.

[Church, 1956] A. Church. *Introduction to Mathematical Logic*, Princeton University Press, Princeton, 1956.

[Copeland, 2002] B. J. Copeland. The genesis of possible worlds semantics, *Journal of Philosophical logic* 31, 99-137, 2002.

[Cresswell, 1975] M. J. Cresswell. Hyperintensional logic, *Studia Logica* 34, 25-38, 1975.

[Cresswell, 1985] M. J. Cresswell. *Structured Meanings: The Semantics of Propositional Attitudes*, MIT Press, Cambridge (Mass.), 1985.

[Divers, 2002] J. Divers. *Possible Worlds*, Routledge, London, 2002.

[Dummett, 1973] M. Dummett. *Frege, the Philosophy of Language*, Duckworth, London, 1973.

[Dummett, 1981] M. Dummett. *The Interpretation of Frege's Philosophy*, Duckworth, London, 1981.

[Etchemendy, 1990] J. Etchemendy. *The Concept of Logical Consequence*, Harvard University Press, Cambridge (Mass.), 1990.

[Frege, 1891] G. Frege. *Function und Begriff (Ein Vortrag, gehlaten inder Sitzung vom 9.1.1891 der Jenaischen Gesellschaft für Medizin und Naturwissenschaft)*, Jena, 1891.

[Frege, 1892a] G. Frege. Über Begriff und Gegenstand, *Vierteljahrschrift für wissentschaftliche Philosophie* 16, 192-205, 1892.

[Frege, 1892b] G. Frege. Über Sinn und Bedeutung, *Zeitschrift für Philosophie und philosophische Kritik* 100, 25-50, 1892.

[Frege, 1893] G. Frege. *Grundgesetze der Arithmetik I-II*, Pohle, Jena, 1893; 1903.

---

[39] For an elaboration of the approach of Tichý in this direction see Materna [1998].

[Frege, 1983] G. Frege. *Nachgelassene Schriften* (ed. H. Hermes et al.), Meiner, Hamburg, 1983.

[Gallin, 1975] D. Gallin. *Intensional and Higher-order Modal Logic*, North-Holland, Amsterdam, 1975.

[Gamut, 1991] L. T. F. Gamut. *Logic, Language and Meaning*, vol. 2, University of Chicago Press, Chicago, 1991.

[Goldblatt, 1975] R. Goldblatt. First-order definability in modal logic, *Journal of Symbolic Logic* 40, 35-40, 1975.

[Goldfarb, 1979] W. D. Goldfarb. Logic in the twenties: the nature of the quantifier, *Journal of Symbolic Logic* 44, 351-368, 1979.

[Henkin, 1950] L. Henkin. Completeness in the theory of types, *Journal of Symbolic Logic* 15, 81-91, 1950.

[Hintikka, 1969] J. Hintikka. *Models for Modalities*, Reidel, Dordrecht, 1969.

[Hintikka, 1978] J. Hintikka. Impossible possible worlds vindicated, in *Game-Theoretical Semantics* (ed. E. Saarinen), Reidel, Dordrecht, 367-379, 1978.

[Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*, Methuen, London, 1968; revised edition *A New Introduction to Modal Logic*, Routledge, London, 1996.

[Janssen, 1997] T. M. V. Janssen. Compositionality, in [van Benthem and ter Meulen, 1997].

[Janssen, 2001] T. M. V. Janssen. Frege, contextuality and compositionality, *Journal of Logic, Language and Information* 10, 115-136, 2001.

[Kripke, 1963] S. Kripke. Semantical analysis of modal logic I (normal modal propositional calculi), *Zeitschift für mathematische Logik und Grundlagen der Mathematik* 9, 67-96, 1963

[Kripke, 1965] S. Kripke. Semantical analysis of modal logic II (non-normal modal propositional calculi), in *The Theory of Models* (ed. J. W. Addison et al.), North-Holland, Amsterdam, 206-220, 1965.

[Lewis, 1932] C. I. Lewis and C. H. Langford. *Symbolic Logic*, Century, New York, 1932.

[Lewis, 1968] D. K. Lewis. Counterpart theory and quantified modal logic, *Journal of Philosophy* 65, 113-126, 1968.

[Lewis, 1972] D. K. Lewis. General semantics, in *Semantics of Natural Language* (ed. D. Davidson & G. Harman), Reidel, Dordrecht, 1972.

[Lewis, 1975] D. K. Lewis. Languages and language, in *Minnesota Studies in the Philosophy of Science* VII (ed. K. Gunderstone), University of Minnesota Press, Minneapolis, 1975.

[Lewis, 1986] D. K. Lewis. *On the Plurality of Worlds*, Blackwell, Oxford, 1986.

[Loux, 1979] M. J. Loux, ed. *The Possible and the Actual: Readings in the Metaphysics of Modality*, Cornell University Press, Ithaca, 1979.

[Materna, 1998] P. Materna. *Concepts and Objects (Acta Philosophica Fennica 63)*, Societas Philosophica Fennica, Helsinky, 1998.

[Mendelson, 1964] E. Mendelson. *Introduction to Mathematical Logic*, Wadsworth & Brooks, Pacific Grove, 1964.

[Montague, 1970a] R. Montague. English as a formal language, in Visenti, B. *et. al.* (eds): *Linguaggi nella Società et nella Tecnica*, Milan, 1970, 189-224; reprinted in [Montague, 1974].

[Montague, 1970b] R. Montague. Universal grammar, *Theoria* 36, 373-398, 1970; reprinted in [Montague, 1974].

[Montague, 1973] R. Montague. The proper treatment of quantification in ordinary English, in *Approaches to Natural Language*, (ed. J. Hintikka et. al.), Reidel, Dordrecht, 221-242, 1973; reprinted in [Montague, 1974].

[Montague, 1974] R. Montague. *Formal Philosophy: Selected Papers of R. Montague* (ed. R. Thomason), Yale University Press, New Haven. 1974.

[Neale, 2001] S. Neale. *Facing Facts*, Clarendon Press, Oxford, 2001.

[Partee and Hendricks, 1997] B. Partee. and H. L. W. Hendricks. Montague grammar, in [van Benthem and ter Meulen, 1997, 5-92].

[Partee, 1982] B. Partee. Belief sentences and limits of semantics, in *Processes, Beliefs and Questions* (eds. S. Peters and E. Saarinen), Reidel, Dordrecht, 1982.

[Partee, 1976] B. Partee, ed. *Montague Grammar*, Academic Press, 1976.

[Pelletier, 2001] F. J. Pelletier. Did Frege believe Frege's principle?, *Journal of Logic, Language and information* 10, 87-114, 2001.

[Peregrin, 2000] J. Peregrin. The "natural" and the "formal", *Journal of Philosophical Logic* 29, 75-101, 2000.

[Peregrin, 2005] J. Peregrin. Is compositionality an empirical concept?, in M. Werning, E. Machery & G. Schurz (eds.): *The Compositionality of Concepts and Meanings*, Ontos, Frankfurt, 135-150, 2005.

[Peregrin, 2006] J. Peregrin. Semantics as based on inference, in *The Age of Alternative Logics* (ed. J. van Benthem et al.), Kluwer, Dordrecht, to appear.

[Quine, 1992] W. V. O. Quine. *Pursuit of Truth* (*revised edition*), Harvard University Press, Cambridge (Mass.), 1992.

[Quine, 1960] W. V. O. Quine. *Word and Object*, MIT Press, Cambridge (Mass.), 1960.

[Shapiro, 1991] S. Shapiro. *Foundations without Foundationalism*, Clarendon Press, Oxford, 1991.

[Tarski, 1936] A. Tarski. Über den Begriff der logischen Folgerung, *Actes du Congrès International de Philosophique Scientifique* 7, 1-11, 1936; English translation 'On the Concept of Logical Consequence' in Tarski: *Logic, Semantics, Metamathematics*, Clarendon Press, Oxford, 1956, pp. 409-420.

[Tichý, 1971] P. Tichý. An approach to intensional analysis, *Nôus* 5, 273-297, 1971; reprinted in [Tichý, 2004].

[Tichý, 1975] P. Tichý. What do we talk about?, *Philosophy of Science* 42, 80-93, 1975; reprinted in [Tichý, 2004].

[Tichý, 1978] P. Tichý. De Dicto and De Re, *Philosophia* 8, 1978; reprinted in [Tichý, 2004].

[Tichý, 1978] P. Tichý. Two kinds of intensional logic, *Epistemologia* 1, 143-164, 1978; reprinted in [Tichý, 2004].

[Tichý, 1986] P. Tichý. Constructions, *Philosophy of Science* 53, 514-534, 1986; reprinted in [Tichý, 2004].

[Tichý, 2004] P. Tichý. *P. Tichý's Collected Papers in Logic and Philosophy* (ed. V. Svoboda, B. Jespersen & C. Cheyne), Filosofia, Prague, 2004.

[Zimmerman, 1989] T. E. Zimmermann. Intensional logic and two-sorted type theory, *The Journal of Symbolic Logic* 54, 65-77. 1989.

LOGICALLY POSSIBLE WORLDS AND COUNTERPART SEMANTICS FOR MODAL LOGIC

Marcus Kracht and Oliver Kutz

> Since to belong and to belong of necessity and to be possible to belong are different (for many things belong, but nevertheless not of necessity, while others neither belong of necessity nor belong at all, but it is possible for them to belong), it is clear that there will also be different deductions of each and that their terms will not be alike: rather, one deduction will be from necessary terms, one from terms which belong, and one from possible terms.
>
> Aristotle: *Prior Analytics, Book I, 8*[1]

INTRODUCTION

Modal logic has a remarkable history indeed. Originally conceived as the logic of necessity and possibility, its philosophical roots go back at least as far as Aristotle and the Stoic Diodorus Cronus.[2] While Diodorus's 'Master Argument' could be considered an early example of modal propositional, more precisely temporal, reasoning,[3] the Stagirite's investigations into modal syllogisms probably constitute the earliest serious essay on combining modality with quantification.[4] Interest into modal distinctions kept very much alive during Mediaeval Scholastics, notably in the form of ontological arguments such as St. Anselm's,[5] and only seems to have

---

[1] English translation by Robin Smith [Aristotle, 1989].

[2] Classic (and encyclopaedic) texts on the early history of logic are Bocheński's [Bocheński, 2002] and William and Martha Kneale's [Kneale and Kneale, 1984].

[3] For the 'Master Argument' compare [White, 1984], as well as, for prominent modern reconstructions, Arthur Prior's [Prior, 1955], Oskar Becker's [Becker, 1960], and Jaakko Hintikka's [Hintikka, 1964].

[4] Modern attempts at interpreting and understanding Aristotle's still puzzling elaborations in Prior Analytics include Łukasiewicz's [Łukasiewicz, 1957] in the 1950ies, [Thom, 1979; Thom, 1996], [Patterson, 1995] and [Schmidt, 2002].

[5] Of which we find variants in the work of, for example, Descartes and Leibniz, and a modern counterpart in Kurt Gödel's ontological argument, a descendant of the argument given by Leibniz, being sketched in higher-order modal logic (second- or third-order depending on the interpretation) and which was only posthumously published in [Gödel, 1995]. For details compare Melvin Fitting's [Fitting, 2002b].

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

slowed down with the dawn of the Renaissance. While Leibniz, being an eminent contributor to the development of modern logic, is usually credited with the introduction of the notion of a 'possible world', the birth of modal logic as a technical discipline of mathematical logic is often considered to be Clarence Irving Lewis's investigations into the paradoxes of material implication in 1918[6], and, in the early thirties, Kurt Gödel's interpretation [Gödel, 1932; Gödel, 1933] of Heyting's version [Heyting, 1930] of Brouwer's intuitionistic propositional logic [Brouwer, 1907; Brouwer, 1908] in terms of the modality 'it is provable'[7]. Despite Quine's early attempts at discrediting the intelligibility of modal discourse [Quine, 1947; Quine, 1953], its technical development went hand in hand with the rise of analytic philosophy, and both its technical sophistication as well as its philosophical significance virtually exploded in the 1970s with the unlikely success of Kripke's possible worlds semantics. In the early 21st century, modal logic is not only an important tool in philosophy and linguistics, it also provides, in its many disguises, key formalisms in artificial intelligence and knowledge representation.

In this contribution, however, we shall not review the history of modal logic, nor do we delve deeply into the discussion of its metaphysics.[8]

Rather, we shall concentrate on the interplay between formal results and analytical thought. As argued already in [Hazen, 1979], formal semantic theories such as Lewisian counterpart theory or Kripkean possible worlds semantics are compatible with a wide spectrum of metaphysical views concerning possible worlds, the nature of objects and possibility, and so on. It is therefore pointless to ask whether or not a given formal semantics is ontologically adequate. Besides, many of these questions quickly take a different turn: they become questions of physical nature whose answers often elude the specialists themselves. That theme will be touched upon when we discuss the nature of time. Our main interest here is, however, different and more modest. We shall simply ask in what ways the formal structures reflect the intuitions that they are claimed to model, and in what ways technical results can be interpreted from a philosophical point of view. The most common semantics of modal logics introduce the notions of a **world**, a **possibility** or a **situation**, and impose further structure by means of more or less complex relations, e.g., by the notion of an **accessibility**.

The nature and existence of 'worlds' has been an intensely debated topic in analytic philosophy and metaphysics. Also, the nature of the accessibility relation has largely been unscrutinised. Although nowadays the idea of a binary relation between worlds is the most popular one, many alternatives have been tried (and are being tried). There is, however, a problem with Kripke-semantics, which is both of technical and of philosophical nature: it is incomplete. There exist different logics with the same class of Kripke-frames. This means that the choice of Kripke-semantics encapsulates metaphysical commitments whose precise nature is unfortunately difficult to explicate. This is unproblematic when the semantics is given beforehand and the logic is derived from it. In all other cases, however, we have committed ourselves to the complete logics from the start without knowing whether that is justified. What is more, criteria for completeness are almost impossible to give (a notable exception is Kit Fine's [Fine, 1975]). The remedy is to restrict the domain of interpretation to admissible sets, but this approach is philosophically unsatisfactory as long as there is no clear account of where these sets originate.

---

[6] C. I. Lewis created his first correct modal calculus **S3** of strict implication in 1920 [Lewis, 1920] (an emendation of a system proposed earlier in his [Lewis, 1918]), compare also [Parry, 1970].

[7] In fact, a similar interpretation can be found even earlier in Orlov's [Orlov, 1928].

[8] But compare [Goldblatt, 2003] for the historical development of mathematical modal logic in the 20th century and [van Benthem *et al.*, 2006] for the state-of-the-art as of 2005, [Copeland, 2002] for the genesis of possible worlds semantics, and [Lindström and Segerberg, 2006] for a discussion of various early systems of quantified modal logic. A good starting point for the reader interested in the metaphysics of modality would be the anthologies [Linsky, 1971] and [Loux, 1979], or John Divers's [Divers, 2002].

Another notion that has been met with suspicion is that of an **object**. One problem with the notion of an object is the problem of **transworld identification**. If there are possible worlds, then which objects in a given world are the same as other objects in another world? The answer depends in part on what we think an object is. Saul Kripke and many with him have assumed that objects are transcendental, to use a rather old-fashioned terminology here. This choice was influenced by a particular school of thought which regards possible worlds as man made, constructed out of the things that we know, namely the objects. The worlds contain more or less the same objects, only the facts about these objects may change. This view faces philosophical problems of its own, and has been challenged on these grounds. Technically, its main disadvantage is that it is highly incomplete. Most propositional logics become incomplete when we move to predicate logic.

A different conception of objects is to view them as entirely world-bound; this is embodied in the semantics of counterpart theory as originally conceived by David Lewis (see [Lewis, 1968]). It can be shown to be more general than Kripke-semantics, but it too carries ontological commitments. The third way is two view objects (in type theoretic terms) as **individual concepts**. This makes them both transcendental and world bound. It can be shown that this semantics (in its generalised form) is actually complete in the most general sense. A fourth way is to abandon the notion of object altogether. This is the semantics advocated for by Shehtman and Skvortsov [Skvortsov and Shehtman, 1993]. In this semantics, a structure is an infinite sequence of frames. These frames combine both the notion of object and world in one; worlds can be identified with the elements at level 0; objects are only approximated to a finite degree at each level. It is possible to construct a frame with worlds and objects from it ([Bauer, 2000]).

PART I

WORLDS WITHOUT OBJECTS

1 BASIC CONCEPTS OF MODAL PROPOSITIONAL LOGIC

The language of **modal propositional logic** consists of a countable set $V = \{p_i : i \in \mathbb{N}\}$ of sentence letters, a set C of propositional constants, the connectives $\top$, $\neg$, and $\wedge$, and a set of so-called **modalities** or **modal operators**. A modality can have an arbitrary finite arity. We consider here only the case of a single operator $\Box$ of arity 1. Many of the philosophical considerations do not depend on that choice. For the technical side see [Kracht and Wolter, 1999]. A **modal logic** is a set L which contains all Boolean tautologies, and is closed under 'modus ponens' (MP): from $\varphi \in L$ and $\varphi \to \psi \in L$ infer $\psi \in L$, and substitution (sub): from $\varphi \in L$ infer $\sigma(\varphi) \in L$. Here, σ is defined from a function from variables to formulae; $\sigma(\varphi)$ is obtained by replacing each occurrence of a propositional variable p_i by some formula $\sigma(p_i)$. (This means that substitution is an admissible rule; a rule is **admissible** in L if the tautologies of L are closed under this rule. (MP) is actually a derived rule, that is, a rule applicable in reasoning from premisses.) L is **classical** if the following rule is admissible.

$$\text{(RE)} \qquad \frac{\varphi \leftrightarrow \chi}{\Box\varphi \leftrightarrow \Box\chi} \tag{1}$$

L is **normal** if it is classical, and additionally

1. $\Box(p_0 \to p_1) \to (\Box p_0 \to \Box p_1) \in L$,

2. and (MN) is admissible in L:

$$\text{(MN)} \qquad \frac{\varphi}{\Box\varphi} \tag{2}$$

A **Kripke-frame** is a triple $\langle F, R, \mathfrak{I}\rangle$, where F is a set, the set of **worlds**, $R \subseteq F^2$ the **accessibility relation**, and $\mathfrak{I} : C \to \wp(F)$ a function mapping propositional constants to subsets of F. In what is to follow, we shall usually assume $C = \varnothing$.[9] A **valuation** is a function $\beta : V \to \wp(F)$ which assigns a set of worlds to each sentence letter. A **pointed Kripke-model** is a triple $\langle \mathfrak{F}, \beta, w\rangle$, where $\mathfrak{F}$ is a (Kripke-) frame, β a valuation, and w a world. **Truth** of formulae in a model is then defined as follows: $\langle \mathfrak{F}, \beta, w\rangle \vDash \top$ always holds, and

---

[9] Whenever propositional constants are left out of the language, a Kripke-frame is just a pair $\langle F, R\rangle$.

$$\begin{array}{lll}\langle \mathfrak{F}, \beta, w\rangle \vDash p_i & :\Leftrightarrow & w \in \beta(p_i); \\ \langle \mathfrak{F}, \beta, w\rangle \vDash c & :\Leftrightarrow & w \in \mathfrak{I}(c); \\ \langle \mathfrak{F}, \beta, w\rangle \vDash \neg\varphi & :\Leftrightarrow & \langle \mathfrak{F}, \beta, w\rangle \nvDash \varphi; \\ \langle \mathfrak{F}, \beta, w\rangle \vDash \varphi \wedge \chi & :\Leftrightarrow & \langle \mathfrak{F}, \beta, w\rangle \vDash \varphi \text{ and } \langle \mathfrak{F}, \beta, w\rangle \vDash \chi; \\ \langle \mathfrak{F}, \beta, w\rangle \vDash \Box\varphi & :\Leftrightarrow & \text{ for all } v \text{ such that } w\ R\ v\text{: } \langle \mathfrak{F}, \beta, v\rangle \vDash \varphi. \end{array} \tag{3}$$

Write $\mathfrak{F} \vDash \varphi$ if $\langle \mathfrak{F}, \beta, w\rangle \vDash \varphi$ for all β and w. Given a class $\mathcal{K}$ of frames, let

$$Th(\mathcal{K}) := \{\varphi : \text{for all } \mathfrak{F} \in \mathcal{K}\ :\ \mathfrak{F} \vDash \varphi\}. \tag{4}$$

$Th(\mathcal{K})$ is called the **logic** of $\mathcal{K}$, for, whenever $\mathcal{K}$ is some class of Kripke-frames, $Th(\mathcal{K})$ is a normal modal logic. Conversely, if L is a normal modal logic, we may define

$$Frm(L) := \{\mathfrak{F} :\ \text{for all } \varphi \in L :\ \mathfrak{F} \vDash \varphi\}. \tag{5}$$

These definitions can be applied to any type of semantics. (In that case, if $\mathcal{X}$ is the class of structures, we index Frm with $\mathcal{X}$ to avoid confusion.) These definitions technically equate a modal logic with its set of axioms (or tautologies), and leaves the set of rules invariant.

2 TRANSLATION INTO CLASSICAL LOGIC

In and of itself, modal logic does not embody the commitment to possible worlds of any sort—rather, the doctrines of modal realism and anti-realism are subject to considerable philosophical debate. For instance, while Charles Chihara's book on modal realism [Chihara, 1998] contains an elaborate attempt at constructing a viable position of modal anti-realism based on the idea to explain modalities not in terms of possible worlds but rather in terms of '*how the world could have been*', the bulk of David Lewis's 'On the Plurality of Worlds' [Lewis, 1986] consists of an extended argument of why we cannot legitimately employ 'possible worlds talk' and enjoy its benefits for instance in the philosophy of language without giving it a realistic interpretation [Holton, 2003].

However, both in technical philosophy and in mathematical logic, possible worlds have become a de facto indispensable tool. It was observed that modal laws can be made plausible by using the notion of a possible world and explicating modal operators using quantification over possible worlds. This led to the well-known **standard translation** of modal logic into classical logic (see [van Benthem, 1983]). By rendering the truth conditions given in (3) into formal logic as well, we arrive at a translation into monadic second-order logic, which runs as follows. For each proposition letter p_i we introduce a monadic second-order predicate letter P_i (which is a variable).

$$\begin{aligned} p_i^\dagger &:= P_i(w) \\ (\neg\varphi)^\dagger &:= \neg\varphi^\dagger \\ (\varphi \wedge \chi)^\dagger &:= \varphi^\dagger \wedge \chi^\dagger \\ (\Box\varphi)^\dagger &:= (\forall v)(w\ R\ v \to \varphi^\dagger(v/w)) \end{aligned} \tag{6}$$

A formula φ is satisfiable in a Kripke-frame $\langle F, R\rangle$ if and only if $\varphi^\dagger$ is satisfiable in $\langle F, R\rangle$ (viewed as a model for second-order predicate logic) iff $(\exists\vec{P})\varphi^\dagger$ holds in it, where $\vec{P}$ contains at least the monadic second-order variables free in $\varphi^\dagger$. Hence $\mathfrak{F} \vDash \varphi$ iff $(\forall\vec{P})\varphi^\dagger$ is valid in $\mathfrak{F}$. Thus frame validity is Π^1_1. If this formula has a first-order equivalent then the propositional logic is complete with respect to Kripke frames, by a result of Kit Fine [Fine, 1975].

Second-order logic adds quantifiers to modal logic that it otherwise does not have. Notice that the formulae on the right hand side of (6) contain only two variables. Thus, the translation is into the two-variable fragment of first-order logic whose satisfiability problem is known to be decidable [Mortimer, 1975]. (Of course, this does not imply the decidability of modal logics involving additional axioms—this is a science on its own, compare [Ohlbach *et al.*, 2001].) From a philosophical point of view this translation raises questions. It can be used to explain modal talk as talk of possible worlds in disguise; this is a reduction from right to left. Modal logic is just talk about possible worlds in a syntactic disguise. It cannot, however, be used to explain talk about possible worlds through modal operators, because the latter is less expressive. There are first-order formulae which have no equivalent in modal logic. This has been observed in [Lewis, 1968] already, noting that this makes modal logic less expressive than it should be if it is to render natural language statements. This point has been picked up by [Hazen, 1976], who extends modal logic by an actuality operator, and further by [Forbes, 1985], who allows to index possible worlds in order to be able to refer back to them. The most expressive language is however that of nominals, introduced by Arthur Prior, see also [Bull, 1968]. A comprehensive study of this language has been made in [Blackburn, 1993]. Nominals are propositional variables that have to be true at precisely one world—they can serve as *names* for worlds. This essentially reifies the worlds inside modal logic and allows the backward pointing devices argued for in [Forbes, 1985] (possibly with the help of the **universal modality**, whose interpretation is basically that of the first-order quantifier over worlds).

3 ONTOLOGY AND DUALITY THEORY

Ludwig Wittgenstein says in the Tractatus that the world is everything which is the case, or, in modern terminology, a complete state of affairs. A view that identifies worlds with complete (or partial) states-of-affairs denies them their own ontological status, a position for which David Lewis coined the term '**linguistic**

ersatzism' [Lewis, 1986].[10] Kripke-frames, on the other hand, treat worlds as primitive objects. Prima facie this seems justified in the case of time points or locations. They provide examples of entities that can in principle serve as worlds in the sense of containers which are inhabited by objects and relations (see [Forbes, 1985]). If worlds exist and are not simply states-of-affairs then the possibility arises of there being two different worlds in which the same facts are true (call them **twin worlds**). Again, whether this is a welcome situation is debated (see below). In duality theory, one studies the construction of frames from algebras and algebras from frames. It turns out that the two model structures are not identical. In passing from frames to algebras, the distinction between twin worlds is lost. On the other hand, the construction of a frame from an algebra may now introduce new worlds to that frame. (They instantiate a complete state-of-affairs that is only finitely satisfied in the original frame.)

A **Boolean algebra with an operator** is a structure of the form $\mathfrak{B} = \langle B, 1, -, \cap, \blacksquare \rangle$, where $\langle B, 1, -, \cap \rangle$ is a Boolean algebra. $\mathfrak{B}$ is a **modal algebra** if in addition

$$\blacksquare 1 = 1, \qquad \blacksquare(a \cap b) = \blacksquare a \cap \blacksquare b \tag{7}$$

Given an assignment $v : V \to A$, there is a unique homomorphism $\overline{v}$ which assigns to every formula over V a value in A. We write $\mathfrak{A} \vDash \varphi$ if $\overline{v}(\varphi) = 1$ for all assignments v and $Th(\mathfrak{A}) := \{\varphi : \mathfrak{A} \vDash \varphi\}$. $Th(\mathfrak{A})$ is classical; it is normal iff $\mathfrak{A}$ is a modal algebra. [Jónsson and Tarski, 1951] have initiated the study of such algebras and shown that they can be embedded in complete Boolean algebras. The atoms of these algebras can be identified with the worlds. Given a logic, modal algebras can be constructed as follows. Let $\mathfrak{T}(V)$ be the algebra of formulae over the set of variables V. Put

$$\varphi \sim_L \chi :\Leftrightarrow \varphi \leftrightarrow \chi \in L \tag{8}$$

L is classical iff $\sim_L$ is a congruence on $\mathfrak{T}(V)$. $\mathfrak{F}_L(V) := \mathfrak{T}(V)$ is an algebra, called the **canonical** or **Lindenbaum-algebra**. It turns out that for countably infinite V,

$$L = \{\varphi : \mathfrak{F}_L(V) \vDash \varphi\} \tag{9}$$

The construction of the algebra uses the language itself. Propositions are formulae modulo equivalence, but there is no assumption on worlds nor states-of-affairs. Propositions form an algebra. From this algebra it is possible to construct worlds as follows. Let $U(\mathfrak{A})$ be the set of ultrafilters of $\mathfrak{A}$. Put $U \ R^* \ V$ if and only if for all $\blacksquare a \in U$ we have $a \in V$. This yields a frame. Unfortunately, the logic of this frame might be different from the logic of the algebra. It therefore becomes necessary to define the following. For $a \in A$ let

$$\widehat{a} := \{U \in U(\mathfrak{A}) : a \in U\} \tag{10}$$

---

[10] Compare also Sider's [Sider, 2002] for a more recent proposal in this direction.

The map $a \mapsto \widehat{a}$ is an isomorphism from $\mathfrak{A}$ onto a subalgebra of the powerset algebra over $\langle U(\mathfrak{A}), R^* \rangle$, with the additional operator

$$\blacksquare a := \{U : \forall V \in U(\mathfrak{A}) : \text{if } U\ R^*\ V \text{ then } V \in a\} \tag{11}$$

The subalgebra can be proper, and can have a different logic. The reason is that the semantics of Kripke-frames is incomplete in the following sense.

Definition 1. Let $\mathcal{L}$ be a class of logics, and $\mathcal{K}$ a class of structures. $\mathcal{K}$ is **complete for** $\mathcal{L}$ if for every logic $L \in \mathcal{L}$, $L = Th(Frm_{\mathcal{K}}(L))$.

If $\mathcal{K}$ is complete for $\mathcal{L}$ then for any two different $L, L' \in \mathcal{L}$ the class of associated structures from $\mathcal{K}$ must be different. Typically, one sets $\mathcal{L}$ to be the extensions of some system (**S4** for example). We shall take here $\mathcal{L}$ to be the class of all normal modal logics. (9) shows that algebraic semantics is complete (for the set of classical logics even). It has been shown in the early 1970s that there are logics which are not completely characterised by their Kripke-frames. In [Thomason, 1972], Thomason has proposed to refine the class of structures as follows. A **general frame** is a triple $\mathfrak{F} = \langle F, R, \mathbb{B} \rangle$, where $\langle F, R \rangle$ is a Kripke-frame and $\mathbb{B} \subseteq \wp(F)$ a system of sets closed under intersection, complement and the operation

$$\blacksquare a := \{v : \text{for all } w\text{: if } v\ R\ w \text{ then } w \in a\} \tag{12}$$

So, put

$$\mathfrak{F}_+ := \langle \mathbb{B}, F, -, \cap, \blacksquare \rangle \tag{13}$$

This is a modal algebra. A subset of F is called **internal** if it is in $\mathbb{B}$. Valuations into general frames may take only internal sets as values. Put

$$\mathfrak{A}^+ := \langle U(\mathfrak{A}), R^*, \{\widehat{a} : a \in A\} \rangle. \tag{14}$$

This is a general frame. A Kripke-frame $\langle F, R \rangle$ can be viewed as the general frame $\langle F, R, \wp(W) \rangle$. From a general frame $\mathfrak{F}$ we defined the algebra $\mathfrak{F}_+$, and from an algebra $\mathfrak{A}$ the general frame $\mathfrak{A}^+$. In general, $(\mathfrak{A}^+)_+ \cong \mathfrak{A}$, but $(\mathfrak{F}_+)^+$ is generally not isomorphic to $\mathfrak{F}$. Completeness with respect to general frames is straightforward to show, using the standard completeness results for algebraic logic and duality. For a logic is always complete for its Lindenbaum algebra. General frames have both worlds and an algebra of internal sets in them. We may view internal sets as propositions. Then, a general frame is **differentiated** if from $v \neq w$ it follows that there is an a such that $v \in a$ and $w \notin a$; it is **tight** if, whenever $\blacksquare a \in v$ also $a \in w$, then $v\ R\ w$; it is **compact** if $\bigcap U \neq \varnothing$ for every ultrafilter of internal sets. $\mathfrak{F}$ is **descriptive** if is differentiated, tight and compact. $\mathfrak{A}^+$ always is descriptive. A world w defines a **state-of-affairs** $H(w)$. v and w are twin worlds if $v \neq w$ and $H(v) = H(w)$. $\mathfrak{F}$ is differentiated iff there are no twin worlds; it is tight if accessibility is definable from possibility; and it is compact if every finitely satisfiable set is contained in some $H(w)$. Together with differentiatedness this means that S is an ultrafilter iff $S = H(w)$ for some w. The internal sets may be seen as the clopen (= closed and open) sets of a topological space; its closed sets are the (possibly infinite) intersections of internal sets. In a descriptive frame, every singleton $\{w\}$ is closed.

3.1 Logical consequence

The semantics is required to be adequate not only in the sense that it handles the tautologies correctly, but that it identifies the deductive rules properly. Again, algebraic semantics takes the lead (see [Wójcicki, 1988], [Czelakowski, 2001] for the general theory of algebraic logic, and [Kracht, 2006] for a survey of modal consequence). A **logical matrix** is a pair $\mathfrak{M} = \langle \mathfrak{A}, D\rangle$ consisting of an algebra of appropriate type and a set $D \subseteq A$, the set of **designated elements**. A valuation is a function v from the set of variables to the domain A; we can also think of it as a homomorphism from the term algebra into $\mathfrak{A}$. φ is **true in $\mathfrak{M}$ under** v if $\overline{v}(\varphi) \in D$. Also, $\Gamma \vDash_{\mathfrak{M}} \varphi$ if every v which makes all formulae of Γ true also makes φ true. $\mathfrak{M}$ is a **matrix for** L if D is closed under the rules of L. Since the only rule of inference is (MP), this comes down to requiring that D is a Boolean filter and that all tautologies map to an element of D under every v. Filters correspond in the dual frame to topologically closed sets of points. We can however specialise to **ultrafilters**, which in the dual frame correspond to single worlds (more exactly, singleton sets). For discussion see [Kracht, 1999].

Theorem 2. For every normal modal propositional logic L there is a class $\mathcal{V}_L$ of matrices such that if $\Gamma \nvdash_L \varphi$ then there exists a $\mathfrak{M} \in \mathcal{V}_L$ such that $\Gamma \nvDash_{\mathfrak{M}} \varphi$.

In fact, we can always choose the matrices to be of the form $\langle \mathfrak{F}_L(V), U\rangle$, U an ultrafilter. This creates the necessary structures from the language itself. Duality theory can be applied to the ontology as follows. Either we look at worlds as primitive objects and define facts as sets of worlds, or we consider facts as primitive and define worlds as maximally consistent sets of facts (see [Brink and Rewitzki, 2002] for a discussion).

3.2 Situations and possibilities

In the philosophical and linguistic literature, often an object quite like a possible worlds is used, namely a *situation*, see [Perry, 1986] and references therein. In contrast to worlds, situations are only partial. Moreover, situations can be 'small' in that their description may require only a single sentence. Situations can be accommodated as follows. Let $\mathfrak{A}$ be a modal algebra. The members of the underlying sets are called **situations**. Given a valuation β and a situation a, write

$$\langle \mathfrak{A}, a, \beta\rangle \vDash \varphi :\Leftrightarrow \beta(\varphi) \geq a \tag{15}$$

Technically, the definitions in [Perry, 1986] amount to taking as matrices only those of the form $\langle \mathfrak{A}, F\rangle$, where $F = \{b : b \geq a\}$ is a principal Boolean filter. If that is the case, however, they are technically just a generalised version of worlds. The use of situations has been defended in [Barwise and Perry, 1983; Barwise and Perry, 1985]. The reasons for pressing for situations against possible worlds has been in addition to the reasons just stated also the claim that possible worlds semantics carries an ontological commitment of their existence. For a defense of possible worlds against this criticism see [Stalnaker, 1986]. Notice that while the

Table 1. Worlds and Situations

entity	algebra	language
situation	filter	deductively closed set
world	ultrafilter	maximally consistent set

above matrices sacrifice bivalence, they do support the same tautologies as the ones based on ultrafilters. Thus, they spell out a supervaluationist account of truth. A related approach is that of [Humberstone, 1981]. A **possibility** is a partial function from propositional letters to truth-values. The interpretation (15) is not truth-functional. Moreover, it is not classical; both φ and $\neg\varphi$ can fail to hold in a. Much has been made of this fact in situation theory. In the present set-up, one can either declare situations to be primitive and construct worlds out of situations, or do the converse. The latter strategy will then treat the truth of a formula in a situation just like supervaluations. Supervaluations are however not compositional. Compositional accounts can of course be given, but they support less formulae than the supervaluation account (for example the law of excluded middle). This is a dilemma also for **quantum logic**. Standardly, quantum logic is supposed to be non-distributive since the algebras are orthomodular lattices. It is however possible to use a supervaluation approach, simply allowing only special filters as sets of designated elements (see [Bigaj, 2001]).

The correspondences are summarised in Table 1. Notice that from the standpoint of duality there is no ontologically primary object. An element of the algebra can be construed either as a set of worlds, or as a set of ultrafilters. Also, an element corresponds to a finitely generated Boolean filter, hence a particular kind of situation. We add that a general frame contains not only worlds and an accessibility relation but also an algebra of sets, whose members we may actually identify as the situations.

4 POSSIBLE WORLDS AS AN ANALYTIC TOOL

Modal logic has proven to be a general tool to model nonclassical logic. This was already apparent in the work of C. I. Lewis from the 1920s [Lewis and Langford, 1932; Parry, 1970]. Lewis' systems **S1**–**S5** were systems of **strict implication**, which were attempts to capture the notion of relevant connection between premiss and conclusion of ordinary language implications. Originally conceived as alternatives to classical logic, it has been observed that one can reduce strict implication to material implication by

$$\varphi \twoheadrightarrow \psi := \Box(\varphi \to \psi) \tag{16}$$

Thus, the properties of $\twoheadrightarrow$ derive straight from the properties of $\Box$. This has fostered the development of modal logic, because the attention turned to modal systems that had a classical background logic rather than to weakenings of classical logic. (Linear logic has brought weakenings back into the game, since the reduction techniques do not work for resource conscious logics.) This method was most successful with the reduction of intuitionistic logic to modal logic. Kurt Gödel gave the following translation from intuitionistic language into monomodal logic [Gödel, 1932; Gödel, 1933].

$$\begin{aligned}\tau(p) &:= \Box p\\ \tau(\varphi \wedge \chi) &:= \tau(\varphi) \wedge \tau(\chi)\\ \tau(\varphi \vee \chi) &:= \Box(\tau(\varphi) \vee \tau(\chi))\\ \tau(\varphi \to \chi) &:= \Box(\tau(\varphi) \to \tau(\chi))\end{aligned} \tag{17}$$

Falsum is mapped to falsum, and $\neg\chi := \chi \to \bot$. This translation reduces intuitionistic logic to modal logic. In the background, however, we have a second reduction, that of modal logic to second-order logic, the standard translation. The possible worlds interpretation that results from this cascaded reduction is actually not dissimilar to the interpretation that Brouwer gave for his logic in the first place; the worlds exemplify stages of knowledge, and it is assumed that knowledge grows as we move to an accessible state (which also happens to be a later stage). Although Brouwer would have opposed a reduction of intuitionism to classical logic (he did oppose Heyting's interpretation), the exactness of the reduction shows that the interpretation is at least viable. Several decades of research have culminated in the following result. Let **Int** be intuitionistic logic. An extension of this logic is of the form $\mathbf{Int} + \Delta$, where Δ is some set of axioms. The system **Grz** is defined by

$$\mathbf{Grz} := \mathbf{S4} \oplus \Box(\Box(p \to \Box p) \to p) \to \Box p \tag{18}$$

Here, $\mathbf{Grz} \oplus \Delta$ denotes the least normal logic containing both **Grz** and Δ.

Theorem 3 (Blok). The map $\mathbf{Int} + \Delta \mapsto \mathbf{Grz} \oplus \{\tau(\delta) : \delta \in \Delta\}$ is an isomorphism from the lattice of intermediate logics onto the lattice of consistent normal extensions of **Grz**.

For details see [Chagrov and Zakharyaschev, 1997]. This much should be enough to demonstrate that modal logic is an exact tool to study intermediate logics. What is more, the original intuitions supplied by Brouwer himself suggest that the formal apparatus actually reflects the analytic notions rather well. Brouwer has spoken about mathematics as a developing structure, where knowledge grows, so that once a statement is accepted, it is accepted for good, while it may be rejected and later come to be accepted. However, its negation is accepted at w only if the formula is not accepted at any v later than w. In intuitionism, facts are always positive; negative facts are in a sense absence of certain positive facts. This asymmetry of positive and negative facts has been removed in the **constructive logic N** of Nelson [Nelson, 1949]. In constructive logic, a formula can be true or false or neither. It is accepted if it is always true and it is rejected if it always false. This is used under the name 'vivid logic' in computer science (see [Pearce and Wagner, 1990; Pearce and Wagner, 1991]). There is a general method to embed intermediate logics into extensions of **N**, which allows to study intermediate logics as special kinds of constructive logics; see [Vakarelov, 1977], and for general results [Sendlewski, 1984; Sendlewski, 1990].

There are also other alternatives to classical logic that cannot be interpreted in terms of modal logic since they are based on the use of the *resources*. This is exemplified in **relevant logic** ([Anderson and Belnap, 1975; Anderson *et al.*, 1992]) and **linear logic** of [Girard, 1987], see also [Troelstra, 1992]. Linear logic, for example, has a connective $\multimap$; $\varphi \multimap \chi$ allows to deduce χ, but the use of this inference means destroying the truth of the antecedent formula, so that it cannot be used again. Actually, it is not impossible to interpret $\multimap$ in classical logic. One can give the translation:

$$\text{if } \varphi \text{ is true at } t \text{ then } \chi \text{ will be true at } u > t.$$

$\multimap$ is hiding the time parameter. It is clear in this interpretation that if φ and $\varphi \multimap \chi$ hold, then the inference to χ means that χ holds at a later time point, but then the truth of φ is not guaranteed any more.

5 ACCESSIBILITY

In the earliest models, which interpreted necessity as logical necessity, the universe was just seen as a set of worlds, with one world singled out as the actual one. However, different interpretations of necessity require to put structure on this set.

5.1 Accessibility relations

For example, if $\Box$ is read as 'it will always be the case that' then the set of worlds must be ordered by temporal precedence. Thus, in addition to worlds we also need an **accessibility relation**. The nature of this relation may be an empirical matter, once the interpretation is decided upon. For example, it is an empirical question how time is structured: it may be discrete, continuous, branching in the future, branching in the past, endless; perhaps we ought to model it as part of relativistic space time rather than just time by itself. These questions are nowadays no longer pure speculation; the structure of space-time is a domain of physics, and thus subject to exact inquiry. Nevertheless, the matter is far from settled. The notion of 'time' continues to be elusive. We may define time either **objectively** or **operationally** ([von Weizsäcker, 1992] prefers the word **phenomenologically**). An objective definition derives the temporal relation from a manifold, typically the space-time continuum, the phenomenological definition instead uses operational criteria in establishing the structure of time. The nature and dimensionality of the space-time manifold is debated. An extreme example is the theory by Hugh Everett ([Everett III, 1956]) according to which the universe develops deterministically by splitting itself into as many worlds as there are possible outcomes for the 'next' moment—the so-called **many worlds interpretation** of quantum mechanics. For such an (objective!) view, probably the most faithful model is based on histories (see §5.3). A more standard approach is to take the n-dimensional Minkowski space. The logic of Minkowski space-time has been identified in all dimensions by Robert Goldblatt in [Goldblatt, 1980] to be **S4.2** for the reflexive relation, and $\mathbf{D4.2} \oplus \Diamond p_1 \wedge \Diamond p_2 \rightarrow \Diamond(\Diamond p_1 \wedge \Diamond p_2)$ [Shapirovsky and Shehtman, 2003] for the irreflexive relation. It follows that, in space-time, the future and past operators do not allow to discriminate different dimensions.

An objective deterministic interpretation is quite compatible with phenomenological nondeterminism. The future may be open simply because we cannot have access to enough facts that will exclude all but one possibility. Many philosophers have been tempted to conclude that this does not apply to the past. Events of the past leave traces ('records') and facts about the past can be answered by looking at these records. However, as [von Weizsäcker, 1985] points out, these records are necessarily partial. Also, the records exist here and now and are used to 'unravel' the present state into a development that has brought it about (see also [Leyton, 1992] on cognitive aspects of this). This unravelling uses the laws of nature, the very laws that we can only deduce by relying on past experience. This creates an abyss for the foundations of physics that has troubled no lesser minds than Einstein himself, see [von Weizsäcker, 1992]. Objectively time may branch into the future while it does not branch into the past; yet, technically it is not easy to explain why that should be so. The only nonsymmetric law is the second law of thermodynamics, which says that order decreases in direction of the future.

On the other hand, notice also that sometimes special interpretations of temporal logic are used which do not depend on the physical nature of time, such as the **temporal logic of programs**. Here, the fact that time is discrete is a design feature of computers.

While temporal interpretations allow the independent study of the accessibility relation, this is not so for the epistemic modalities. Belief worlds cannot be claimed to exist in the same way as future worlds, and their structure is most likely derived from the belief structure of the agent(s). If A believes δ in v then every belief-alternative w must make δ true. Here, the belief in δ comes first, and the accessibility is constructed. Perhaps this is the reason why epistemic modalities generally fail most common logical laws, in contrast to temporal modalities. Indeed, it has been argued among others by Quine that there is a difference between temporal modalities and epistemic modalities [Quine, 1976]. While it makes sense to attribute de re in temporal logic this is meaningless for belief because we lack a notion of sameness of objects across worlds.

5.2 Neighbourhoods

There are alternatives to accessibility relations and the standard conception of truth. [Segerberg, 1971] introduces the notion of a **neighbourhood frame**, which is a pair $\langle W, N\rangle$, where W is a set of worlds and N a function from W to $\wp(\wp(W))$. As before, valuations are functions $\beta : V \rightarrow \wp(W)$, and formulae are true in a frame at a world relative to a valuation.

$$\langle\langle W, N\rangle, \beta, w\rangle \vDash \Box\varphi :\Leftrightarrow \{v : \langle\langle W, N\rangle, \beta, v\rangle \vDash \varphi\} \in N(w) \qquad (19)$$

Neighbourhood frames are also called **Scott-Montague** frames, after [Scott, 1970] and [Montague, 1968]. The theory of a class of neighbourhood frames is a classical modal logic. Evidently, a Kripke-frame can be turned into a neighbourhood frame (put $N(w) := \{S : S \supseteq \{v : w\ R\ v\}\}$). The converse is not true, however. [Gerson, 1975b] has provided examples of neighbourhood-incomplete logics, and [Gerson, 1975a] has shown that there are logics containing **S4** which are determined by their neighbourhood frames but not by their Kripke-frames.

The **logic of conditionals** is an example of a logic for which standard relational semantics is inadequate. The semantics for **counterfactuals** is therefore more complex. [Stalnaker and Thomason, 1970] for example use choice functions, [Gabbay, 1972] introduces a relation that depends also on two propositions ([van Fraassen, 1974] argues that conditional logic has 'hidden variables' in the sense argued for in quantum mechanics).

David Lewis defines structures in which every world has a set of sets of worlds (set of 'spheres') around it [Lewis, 1973]. The structures are special neighbourhood frames, for it is assumed that the set of spheres around w are concentric; that is, if S and T are spheres, then either $S \subseteq T$ or $T \subseteq S$. Worlds within the same sphere of w are thought of to be similar to the same degree to w. Without having to specify numerical values for how similar one world is to the next, it is possible to define that a world v is more similar to w than a world v'. This is the case if all spheres around w which contain v also contain v'. Lewis also considers alternative structures, for example with a comparative similarity relation, but shows that they are interreducible [Lewis, 1971]. Now, $\varphi > \psi$ is taken to be true at w if in all worlds u most similar to w that make φ true, also ψ is true.

The notion of **similarity** between worlds is widely used. For instance, an idea developed in [Suzuki, 1997] is to relativise accessibility by attaching the relations with fuzzy numbers, thus talking about 'degrees of accessibility', an idea that eventually turned into logics where accessibility is regimented by metric spaces or similarity measures, compare e.g. [Kutz *et al.*, 2003]. Quite often, one imagines a world that is as similar as possible to ours, while something specific is different. So we consider worlds in which φ is true, despite the fact that it is not. The spheres around the worlds are assumed by Lewis to be given. Epistemologically, one would like to be able to say how to find this structure, that is, to construe it from properties of the worlds themselves. A primitive solution is that the spheres around w are the worlds of Hamming-distance $\leq n$ for every given n. (This means that at most n values of the primitive letters can be changed in going from w to a world in that sphere.) We could also say that the worlds in the spheres result in the revision of the theory at w. Revision has been studied extensively in recent decades, but this topic is outside the scope of this article.[11]

---

[11] But compare [Alchourrón *et al.*, 1985] for the proposal now referred to as the AGM theory of revision (named after its inventors), and [Lindström and Segerberg, 2006] for an overview.

A related phenomenon that requires more than basic modal logic is the **imperfective paradox**, exemplified in ‘The dog was crossing the road when it got hit by a car.’ If the dog was hit by a car how can we sensibly attribute to the dog that it was about to cross the road? David Dowty, in an attempt to explain the semantics behind the imperfective paradox, takes recourse to **inertia worlds** [Dowty, 1977]. These are worlds which are very much like ours, in particular normal laws of physics and everyday life hold. Angelika Kratzer develops another idea, whereby the actual world lumps propositions together such that they cannot be given up independently [Kratzer, 1989]. Kratzer makes this possible by introducing situations and a refinement relation, the maximal members of which are the worlds. All these are attempts at defining accessibility from notions that are verifiable hic-et-nunc. They are only gradually different from causality in the sense that all of these notions are derived from directly observed facts of the world. Causality is different only inasmuch as it is believed to be the source of the regularity, not the other way around.

The clause (19) contains two quantifiers, an existential and a universal. Moreover, the function N is type-theoretically second-order, therefore allowing for greater expressivity. The idea of modal operators with complex interpretive clauses of any level has been pursued by Mark Brown, with applications to various interpretations (group knowledge, ability, among other, see [Brown, 1988]). His interpretations are different from the one above, though. It can be shown that modal logic of the ordinary sort can interpret these complex ones [Kracht and Wolter, 1999].

5.3 *Histories*

It has been argued that propositions are not true at a world simpliciter but only at a history. This conception, called **Ockhamist**, has been revitalised by Arthur Prior in [Prior, 1967] and formalised in [Burgess, 1979]. Call a **history** a maximal set of worlds linearly ordered by time (this definition presupposes a particular structure of the frame to begin with). A formula is evaluated in a frame at a pair $\langle h, v\rangle$ where h is a history and $v \in h$. Then

$$\begin{aligned}&\langle\langle W, R\rangle, h, v, \beta\rangle \vDash [F]\varphi \\ &\quad :\Leftrightarrow \text{ for all } w \in h \text{ such that } v\; R\; w : \langle\langle W, R\rangle, h, w, \beta\rangle \vDash \varphi\end{aligned} \tag{20}$$

$$\begin{aligned}&\langle\langle W, R\rangle, h, v, \beta\rangle \vDash [P]\varphi \\ &\quad :\Leftrightarrow \text{ for all } w \in h \text{ such that } w\; R\; v : \langle\langle W, R\rangle, h, w, \beta\rangle \vDash \varphi\end{aligned} \tag{21}$$

$$\begin{aligned}&\langle\langle W, R\rangle, h, v, \beta\rangle \vDash \Box\varphi \\ &\quad :\Leftrightarrow \text{ for all } h' \text{ such that } v \in h' : \langle\langle W, R\rangle, h', v, \beta\rangle \vDash \varphi\end{aligned} \tag{22}$$

One may additionally complicate these frames by allowing only certain bundles of histories to be values of propositions. In computer science these systems have received growing attention. It has been shown only fairly recently that Ockhamist temporal logic is more expressive [Reynolds, 2002] and complete axiomatisations have been given among others in [Zanardo, 1985] for discrete time and (for real lines and branching in the future) in [Reynolds, 2003]. A different interpretation, the **Peircean** interpretation (see [Burgess, 1978]) accepts a future statement as true if and only if for for every history through the given world, it will be true at a later point. This means that φ is certain to happen in the future. This can be rendered in an Ockhamist interpretation as $\Box\langle F\rangle\varphi$.

PART II

THE WORLD OF OBJECTS

Combining modality with quantification is a subtle affair. According to Alonzo Church [Church, 1951], the year 1946 should be considered to be the birth of the discipline as a modern logical enterprise, with Ruth C. Barcan publishing 'A Functional Calculus of First Order Based on Strict Implication' [Barcan, 1946] and, almost simultaneously, Rudolf Carnap publishing 'Modalities and Quantification' [Carnap, 1946], these two papers being the first modern systematic treatments mixing *unrestricted* quantification as introduced to logic by Gottlob Frege with modalities, taking up a subject that Aristotle left more than two millennia ago.

Barcan's paper also introduced the now famous Barcan formulae which mix **de re** modalities attributing a necessity to a *thing* (as in 'Everything necessarily exists') with **de dicto** modalities, attributing necessity to *what is said* (as in 'Necessarily, everything exists'); a distinction which essentially goes back to Aristotle's distinction between *composition* and *division* [Parsons, 1995].

6 MODAL PREDICATE LOGIC

The languages of modal predicate logic that we shall consider first differ from the language of modal propositional logic as follows. First, there are neither propositional variables nor propositional constants. Second, we shall distinguish between **first-order** and **second-order** languages of modal predicate logic which differ syntactically as well as with respect to the **substitution principles** assumed, and which are defined thus. In the first-order case, we have a set $U = \{x_i : i \in \mathbb{N}\}$ of **object variables** (we postpone the discussion of object constants to § 7) and, instead of the propositional variables, a set of **predicate constant letters** together with a map Ω assigning to each letter P its **arity** $\Omega(P)$. Though it is possible to add function symbols, we shall not do so, as they are technically eliminable. In that case, terms are just variables. If $\Omega(P) = n$, and the t_i, $1 \leq i \leq n$, are terms, then $P(t_1, \ldots, t_n)$ is a proposition. A set L is a **normal first-order modal predicate logic** (**first-order MPL**) if

1. L contains all instances of the axioms of predicate logic.
2. L contains the first-order instantiations of the tautologies of **K**.
3. L is closed under (MP) and (MN).

Since L comprises standard predicate logic, it is closed under the usual substitution of terms for variables; call this closure under **first-order substitution**. For the treatment of equality, see below, §6.3.

In the second-order case, we replace the propositional variables with sets $\{P_i^n : i \in \mathbb{N}\}$ of **predicate variables** of arity n for each n. A set L of formulae in this language, then, is called a **normal second-order modal predicate logic** if it is closed under substitution of any formula $\varphi(y_1, \cdots, y_n)$ for an occurrence of $P_i^n(x_1, \cdots, x_n)$ in a formula χ; call this closure under **second-order substitution**.[12]

Notice that the second-order MPLs thus defined are not truly second-order since there are no second-order quantifiers. In general, the terms 'modal predicate logic' or 'first-order modal logic' are typically taken to mean second-order modal predicate logic in the sense above, for instance in [Hughes and Cresswell, 1996] and [Fitting and Mendelsohn, 1998].[13]

Much of the previous discussion of modal propositional logic transfers to the predicate setting. It does, however, introduce complications of its own. In what is to follow we shall review some of the proposals for semantics of modal predicate logic that have been made. The standard Kripke-style semantics can obviously never be complete for the reason that already not all propositional logics are complete. The best one can hope for is therefore that a modal predicate logic is complete on condition that its propositional counterpart is. Below, we shall exhibit such a semantics.

While the notion of a predicate is generally taken to be unproblematic, the notion of an object has been the subject of considerable controversy. Part of the controversy can be retraced in type theory as follows. In addition to the type t of truth values and the type of **individuals** e (for predicate logic) we introduce a new type, s, that of possible worlds. Propositions have type $s \to t$. The question is: what type do we associate with object variables? Kripke assumes they have type e simpliciter; others have argued they should be seen as having type $s \to e$ (**individual concepts**). In the latter case, predicates may either be regarded as relations of individuals or as relations of individuals under a concept. In the first case, a binary predicate letter is interpreted as $s \to (e \to (e \to t))$, in the second case as $s \to ((s \to e) \to ((s \to e) \to t))$. Once the types are fixed, one can actually infer from Henkin's completeness proof for **Simple Type Theory** [Henkin, 1950] a general completeness theorem for various semantics. However, the resulting structures are rather unintuitive.

---

[12] This substitution must be formulated with care to prevent accidental capture of the y_i. The best way to do this is to rename any bound occurrence of y_i, $1 \leq i \leq n$, in χ prior to substitution. For a detailed discussion consult [Kracht and Kutz, 2005].

[13] Other types of first- or higher-order intensional logics that we will not discuss here include Alonzo Church's attempt to formalise Frege's logic of sense and denotation [Church, 1951], and, growing out of this, the work of Richard Montague [Montague, 1969], Daniel Gallin [Gallin, 1975], and Melvin Fitting [Fitting, 2002b]; compare [Lindström and Segerberg, 2006] and [Braüner and Ghilardi, 2006] for more details.

6.1 The classical view

Models for modal predicate logic are defined by Saul Kripke as follows. A frame is once again a pair $\langle F, R\rangle$, with F a set of worlds, and $R \subseteq F^2$. Each world is however now a pair $w = \langle D_w, \mathfrak{I}_w\rangle$, where D_w is a set (the domain of w) and $\mathfrak{I}_w$ an interpretation of the predicates and constants in D_w. If $w\ R\ v$ then $D_w \subseteq D_v$. This means every valuation into D_w is a valuation into D_v as well. This allows for a straightforward definition of the truth of $\Box\varphi$:

$$\langle\langle F, R\rangle, \beta, w\rangle \vDash \Box\varphi :\Leftrightarrow \text{ for all } v \rhd w : \langle\langle F, R\rangle, \beta, v\rangle \vDash \varphi \tag{23}$$

The condition on growing domains corresponds to the validity of the **converse Barcan formula** $\Box\forall x\varphi(x) \to \forall x\Box\varphi(x)$. The converse of this implication, the **Barcan formula**, $\forall x\Box\varphi(x) \to \Box\forall x\varphi(x)$, is not generally valid.[14] It is, however, if we have $D_v = D_w$ for all $v\ R\ w$ so that domains are effectively constant. Another principle generally true in these frames is the **necessity of identity**

$$(x = y) \to \Box(x = y) \tag{24}$$

which is defended by Saul Kripke in the context of arguing for his **causal theory of reference** [Kripke, 1972].[15]

It follows from the strong version of Leibniz' Principle (see §6.3 below):

$$s = t \to (\varphi(s/x) \leftrightarrow \varphi(t/x)) \tag{25}$$

In this interpretation, objects are transcendental, and so the question of trans-world identity does not arise. This fact is responsible for the necessity both of identity and nonidentity. Also, domains have to grow; objects cannot be removed once they have been introduced.

The logic of these structures, **QK**, is characterised over the languages introduced above by the following additional axioms.

1. The converse Barcan formula.

---

[14] This appeared first in 1946 as Axiom 11 in Ruth C. Barcan's [Barcan, 1946], although as the notational variant $\Diamond\exists x\varphi(x) \twoheadrightarrow \exists x\Diamond\varphi(x)$ involving the existential diamond $\Diamond$ and strict implication.

[15] Which is closely related to Ruth Barcan Marcus' theory of 'contentless directly referential tags' account of proper names developed in her [Marcus, 1961], a famous discussion of which appeared in [Marcus *et al.*, 1962], listing R. B Marcus, W. V. Quine, S. Kripke, J. McCarthy, and D. Føllesdal as discussants. For a historical account, compare [Smith, 1998].

2. $x = y \to \Box(x = y)$.
3. $x \neq y \to \Box(x \neq y)$.

If a strengthening is obtained by adding only purely modal propositional axioms, we refer to it as **quantified** L, where L is the modal counterpart. Often, the symbol $\mathbf{Q}L$ is used. Other strengthenings involve interactions between quantifiers, identity, and modal operators. These are the most popular logics under investigation. The remaining option, to just strengthen the predicate logical axioms, is rarely considered, partly because these axioms can easily be eliminated by conditionalising the other axioms.

6.2 Free logic

There are many problems with the converse Barcan-formula. Obviously, objects that exist in this world need not exist in others. One way to fix this is to introduce **free logic** (compare [Lambert, 1997] for an overview). Free logic adds to predicate logic an existence predicate, denoted here by E, which is interpreted like an ordinary 1-place predicate. The set $\mathfrak{I}(E)$ is called the **domain of existence**, and its members are said to **exist**. Objects outside of $\mathfrak{I}(E)$ exist only in a weaker sense. The standard quantifiers range only over existing objects. Therefore, some laws of predicate logic, notably the law of **universal instantiation**, have to be weakened.[16] The axiomatisation is as follows. In addition to the axioms and rules of **K** we assume (where $fv(\varphi)$ denotes the variables free in φ)

1. If $x \notin fv(\varphi)$, then $\varphi \to (\forall x)\varphi$ is a tautology.
2. (a) $(\forall x)E(x)$.
   (b) If y is free for x in $\varphi(x, \vec{z})$ and $y \notin \vec{z}$, then $(\forall x)\varphi(x, \vec{z}) \to (E(y) \to \varphi(y, \vec{z}))$ is a tautology.
3. The tautologies are closed under (UG): $\varphi/(\forall x)\varphi$.

If equality is a symbol of the language, we add the following.

1. $x = x$ is a tautology.
2. $x = y \to (\varphi(x) \to \varphi(y//x))$ is a tautology, where $\varphi(y//x)$ denotes the result of replacing one or more free occurrences of x by y, where y does not become accidentally bound in that occurrence.

The last is the strong Leibniz' Principle, see the discussion below. [Hintikka, 1969] has observed that existence can be defined by

$$E(x) \leftrightarrow (\exists y)(y = x) \tag{26}$$

---

[16] Historically, this weakening of the rule of universal instantiation goes back to the work of Saul Kripke [Kripke, 1963] and Karel Lambert [Lambert, 1963].

Alternatively, one can introduce quantifiers $\bigwedge$ and $\bigvee$, which range over the entire universe, and define

$$(\forall x)\varphi := \bigwedge_x E(x) \to \varphi \tag{27}$$

(See for example [Hodes, 1986].) The difference between these two types of quantifiers is often labelled **possibilist** ($\bigwedge$) versus **actualist** ($\forall$), *cf.* [Prior, 1957; Hughes and Cresswell, 1996].

Using free logic one can eliminate a problem that besets the standard semantics. If domains have to grow, there is no way in which some object may cease to exist at some point. However, we do want to say that Aristotle does not exist now, even though at some point he did. Rather than trying to accommodate for this using different model structures, it seems better to use free logic. (The advice to use free logic in the context of modal logic was explicitly given in [Scott, 1970], compare also [Garson, 1991].) Once free logic is introduced we may even assume that $D_w = D_v$ if $w\ R\ v$, so that one may effectively assume that all models have the same domain. However, notice that free logic is more flexible; neither the Barcan formula nor its converse hold any more because objects can freely pass in and out of existence. The converse Barcan formula holds only if $\mathfrak{I}_w(E) \subseteq \mathfrak{I}_v(E)$ for $w\ R\ v$.

There are two ways to go from here. [Forbes, 1985] allows predicates to take as values in a given world any set of n-tuples drawn from the domain, regardless of their existence, thus contradicting the **Falsehood Principle** of Kit Fine [Fine, 1981] which states that the extension of a predicate at a given world is a set of existing n-tuples. This principle can be reinstated in the form of the axioms

$$\bigwedge_{x_1} \cdots \bigwedge_{x_n} (P(x_1, \ldots, x_n) \to E(x_1) \wedge \cdots \wedge E(x_n)) \tag{28}$$

for every primitive n-ary predicate letter P.

It should be clear, however, that the semantical problems related to nonexistents or fictional discourse in general (or talk of ‘impossible’ objects such as the infamous ‘round square’ for that matter) are rather complex and not simply solved by moving to free logic, as witnessed, for instance, by Gideon Rosen’s ‘Modal Fictionalism’ [Rosen, 1990] (which builds on David Lewis’s counterpart theory while trying to avoid its ontological commitments) or Terence Parsons’s ‘Nonexistent Objects’ [Parsons, 1980].

6.3 Identity, substitution and Leibniz’ Law

Leibniz’ Law asserts that identicals are substitutable ‘salva veritate’. In modal predicate logics, this applies both to propositions and to objects. We shall look at propositions first. Standardly, Leibniz’ Principle is rendered as follows, which we call the **strong Leibniz’ Law**:

$$p \leftrightarrow q \vDash_L \varphi(p/r) \leftrightarrow \varphi(q/r) \tag{29}$$

Boolean logic satisfies this. A modal logic satisfies (29) iff it satisfies

$$p \leftrightarrow q \vDash \Box p \leftrightarrow \Box q \tag{30}$$

Since the only rule of inference is (MP), this is equivalent to $p \to \Box p \in L$. The notion is therefore trivialised. [Kracht, 2004] argues that Leibniz' Principle is not about accidental truth but about meaning. From this perspective, the principle actually amounts to the admissibility of the rule (RE), repeated here for convenience.

$$\frac{\varphi \leftrightarrow \chi}{\Box\varphi \leftrightarrow \Box\chi} \tag{31}$$

Leibniz' Law in this form is valid in every classical, and hence in every normal modal logic (as opposed to the strengthened version (29), which is valid only if $p \to \Box p \in L$).

When talking about identity and Leibniz' law with respect to objects, typically Leibniz' Principle is spelled out in the strong version. The **weak** version is this.

$$\frac{s = t}{\varphi(s/x) \leftrightarrow \varphi(t/x)} \tag{32}$$

Notice that such a substitution principle is actually derivable in the case of classical first-order logic and more generally for any logic that is axiomatised by *unrestricted* schemata. Nevertheless, by assuming unrestricted second-order substitution for a given logic L one automatically extends the underlying modal theory of identity. *E.g.*, given that $(x = y) \to (P(x, x) \to P(x, y))$ is an admissible instance of Leibniz' Law, second-order substitution yields $(x = y) \to (\Box(x = x) \to \Box(x = y))$ and hence $(x = y) \to \Box(x = y)$. Actually, this situation is one of the reasons for introducing a weaker base logic than the usual **QK**. [Kracht and Kutz, 2002] worked with a system called **FK** which is a combination of propositional modal logic **K** and **positive free logic**, **PFL**.[17]

If equality is introduced, the base logic is enriched by a weak form of Leibniz' Law, which we called the **Modal Leibniz' Law**. This basically results from the usual Leibniz' Law by restricting the Quinean principle of the 'substitutability of identicals' to those instances that do not entail 'transworld-identifications' of individuals of any kind. Briefly, if $x = y$ and the variable x appears free within the scope of a modal operator, then either *all* or *no* occurrence of x may be replaced by y. Hence, $(x = y) \to (\Box(x = x) \to \Box(x = y))$ is not an admissible substitution, which blocks the provability of the necessity of identity.[18]

A natural solution to the above problem of generating possibly unintended theorems involving equality is therefore to deal with second-order logics without identity and to add a modal theory of identity, or, alternatively, to incorporate the theory of identity into the logic while restricting substitution in an appropriate way.

---

[17] For a more detailed discussion and an argumentation why free logic is not only useful but necessary, *cf.* [Kracht and Kutz, 2002] or [Kutz, 2000].

[18] Thus, the modal operators behave quite similar to what is known as an *unselective binder* in linguistics.

Since closure under second-order substitutions has a quite different flavour in a propositional as opposed to a predicate logic setting, there are a number of reasons to be interested in first-order MPLs and to treat them as *genuine* logics. We list just a few of them. First, if atomic propositions/predicates enjoy a special status—like in certain logics of time—then substitution of complex formulae for atoms may not be admissible. Actually, this was one of the reasons for Robert Goldblatt to introduce a similar distinction in the propositional case and to call it a 'significant conceptual change' (compare his [Goldblatt, 1992]). Similarly for the case where basic predicates may be *intensional*. Second, if one works with a weak *logic of identity*, then a restriction of substitution is unavoidable. Last but not least, if generalised semantics are considered, there are naturally defined frame classes whose logic is only first-order closed. However, one can also argue in favour of closure under second-order substitution as a *defining property* of the general concept of a 'logic', which has been attempted for the case of MPL in [Bauer and Wansing, 2002].[19]

To make this discussion less theoretical, let us have a brief look at arguably the most prominent logic for which general substitution fails: Rudolf Carnap's logic of **logical truth** (L-truth) [Carnap, 1946; Carnap, 1947]. Let us call the propositional version of Carnap's logic **C**, following Schurz [2001] where a detailed discussion can be found. **C** as well as the quantified version **QC** of Carnap's logic are often mistaken to be versions of (quantified) **S5**, but they are in fact much stronger and have rather unusual logical properties.

The source of the differences between **S5** and **C** is, however, easily located. While Carnap considers a sentence to be logically true if it is valid in a *fixed* set $\mathbf{W}$ of all possible interpretations of the language, in Kripke's account [Kripke, 1959], a sentence is logically true if it is valid in each variable subspace $V \subset \mathbf{W}$.[20] Thus, while in Carnap's logic we have for every satisfiable sentence A the logical truth $\Diamond A$ as a theorem of **C**, Kripke's **S5** does not contain any nontrivial possibility theorems at all. Whenever $\Diamond A$ is in **S5**, so is A. And exactly these nontrivial possibility theorems of **C** account for the fact that substitution fails in general, as should be obvious.[21] As [Schurz, 2001] argues, if one identifies necessity with **logical necessity** then **C** is the *only* complete modal logic, compare also [Cocchiarella, 2002].

In fact, as Carnap [1946] and Thomason [1973] independently showed, **S5** corresponds exactly to the substitutionally closed fragment of **C**.

---

[19] According to Alfred Tarski [Tarski, 1936], the idea to define *logical truth* syntactically as the requirement of closure under all substitution instances goes back to Bernhard Bolzano, compare also [Schurz, 1998].

[20] This was further analysed by Nino Cocchiarella [Cocchiarella, 1975] with his distinction between *primary* and *secondary* semantics. Notice also the analogy to the frame/general frame distinction.

[21] Clearly, $\bot$ cannot be substituted for A in $\Diamond A$. It also follows that, unlike **QS5**, the theorems of **QC** are not even recursively enumerable.

6.4 Completion

Propositional modal logics are not always complete with respect to Kripke-semantics. The same holds a fortiori for modal predicate logics. There is a general procedure for turning an algebraic model for the propositional language into a model for the predicate language, which goes back to [Lawvere, 1969], and ultimately to the completeness proof by Henkin for **Simple Type Theory**. This will be discussed in § 11.

There is an idea analogous to the construction of a general frame: we start with a Kripke-frame and restrict the interpretation of predicates. This is a purely linguistic restriction; there is nothing in the ontology itself that suggests why such a restriction is warranted. The technical details are as follows. We introduce a 'tower' of sets that represent the possible values for predicates. Before we can give a precise definition, some technical preliminaries are necessary. If $\sigma : \{1, \ldots, m\} \to \{1, \ldots, n\}$ is a map, and $\vec{a}$ is an n-tuple, put $\sigma(\vec{a}) := \langle a_{\sigma(1)}, \ldots, a_{\sigma(m)} \rangle$. This defines a map from n-tuples to m-tuples; it is lifted to sets as follows.

$$\sigma(c) := \{\sigma(\vec{a}) : \vec{a} \in c\} \tag{33}$$

Also, the **cylindrification** E_m for $m \leq n$ is defined by

$$E_m(c) := \{\langle a_1, \ldots, a_n \rangle : \exists d. \langle a_1, \ldots, a_{m-1}, d, a_{m+1}, \ldots, a_n \rangle \in c\} \tag{34}$$

A **generalised Kripke-frame** is a structure $\langle W, R, \mathcal{C} \rangle$, where $\langle W, R \rangle$ is a Kripke-frame and $\mathcal{C} = \{\mathbb{C}_n : n \in \mathbb{N}\}$ is a sequence of sets $\mathbb{C}_n \subseteq \wp(D^n)$ satisfying the following postulates.

1. $\mathbb{C}_n$ is a modal algebra for every n.
2. If $c \in \mathbb{C}_n$ and $\sigma : m \to n$ then $\sigma(c) \in \mathbb{C}_m$.
3. If $c \in \mathbb{C}_n$ and $n \geq m$ then $E_m(c) \in \mathbb{C}_n$.

We call such a system a **tower**. Now, the condition is that for a predicate P, the interpretation $\mathfrak{I}_w(P)$ is a member of $\mathbb{C}_n$. The following clause is added.

$$\langle W, R, \mathcal{C}, \beta, w \rangle \vDash P(x_{\sigma(1)}, \ldots, x_{\sigma(m)}) \Leftrightarrow \beta(\langle x_1, \ldots, x_m \rangle) \in \sigma(\mathfrak{I}_w(P)) \tag{35}$$

By assumption, the interpretation of a formula is a member of the tower. The sets of the tower do not change from world to world.

So far, we have discussed the syntactic definition of modal predicate logics, some of the standard semantical approaches including Kripkean possible worlds semantics and free logic, as well as related problems, as for instance the validity of the Barcan formulae in standard semantics and the problems with substitution and identity. Many of the more general semantics that have been proposed are designed to deal with one or the other of these problems, with the eventual goal of very general completeness theorems encapsulating a minimum of ontological

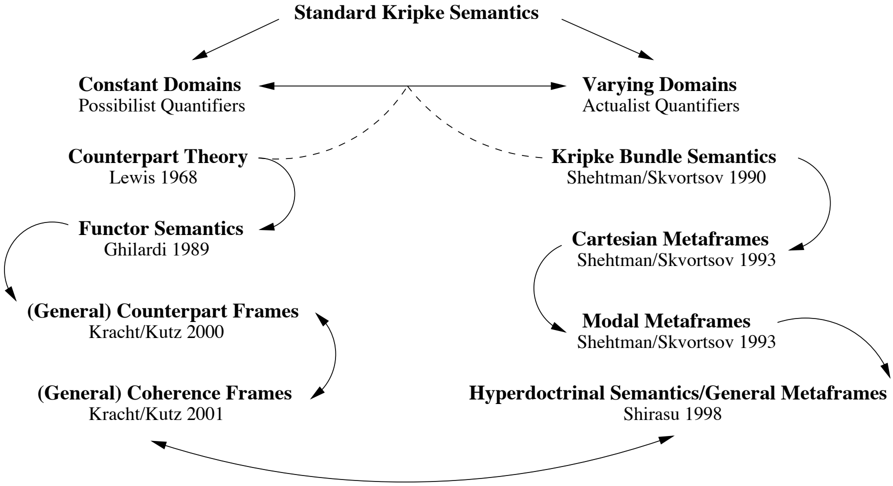


Figure 1. An overview of various semantics for modal predicate logics ordered by relative generality.

presuppositions, or—as Johan van Benthem has formulated it—referring to the merits of completeness proofs in general, that 'embody conditions of adequacy on empirical theories in semantics' [van Benthem, 1986].

The term **Kripke-type semantics**, as introduced in [Skvortsov and Shehtman, 1993], refers to a family of semantics that, rather than introducing purely algebraic structures, tries to keep as much as possible of the basic intuitions of 'possible worlds semantics' (the 'geometric approach') while transcending its scope [Kutz, 2003].

The main differences to the standard semantics are twofold: first, instead of taking a Kripke frame, that is, a set of possible worlds together with an accessibility relation, and to enrich it by assigning domains to worlds, one starts with a family of first-order domains and adds some set of functions or relations between the domains, which in turn *define* accessibility between worlds. Hence, a *plain* accessibility relation is no longer a primitive of the frame but rather depends on (can be defined by) the functions/relations being present. This leads to the second fundamental difference, namely that there may indeed be many distinct functions/relations between two given worlds, a feature that we will label 'modes of transgression' in the context of counterpart semantics that we will investigate in the next section. Figure 1 shows some of the different proposed Kripke-type semantics and their interdependencies. An arrow from A to B indicates that the semantics A is a special case of (can be simulated by) semantics B.

7 COUNTERPART SEMANTICS.

David Lewis [Lewis, 1968] provided the first formal theory of counterparts. It is a two-sorted first-order theory, whose sorts are objects and worlds, and which has four predicates: $W(x)$ says that x is a world, $I(x, y)$ that x is in the world y, $A(x)$ that x is an actual object, and $C(x, y)$ that x is a counterpart of y. The postulates codify that every object is in one and only one world, that counterparts of objects are objects, that no two different objects of the same world can be counterparts of each other, any object is a counterpart of itself, and that there is a world that contains all and only the actual objects and which is nonempty.

The connection to the standard modal language is obtained by translating sentences of MPL to the counterpart theory via the so-called **Lewis translation** (for details compare [Lewis, 1968]).

For instance, an expression of the form $[\Diamond\varphi(t_1, \ldots, t_n)]^v$ ('$\Diamond\varphi(t_1, \ldots, t_n)$ holds in world v') is recursively translated to

$$\begin{aligned}&\exists w \exists s_1 \ldots \exists s_n (W(w) \wedge I(s_1, w) \wedge C(s_1, t_1) \wedge \ldots \\ &\qquad \ldots \wedge I(s_n, w) \wedge C(s_n, t_n) \wedge [\varphi(s_1, \ldots, s_n]^w) \qquad (36)\end{aligned}$$

However, there are several problems related to this translation approach. First, in its original form, counterpart theory does not validate all theorems of the smallest quantified modal logic **QK** since it fails to generally support the principle of **Box distribution**

$$\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi),$$

which is valid in all normal modal logics, compare [Hughes and Cresswell, 1996]. In fact, the simultaneous quantification over both worlds and individuals in counterpart theory obscures the notion of accessibility between worlds and thus leads to the semantic refutability of certain **K**-theorems [Kutz, 2000].

Further, it has often been argued that the standard modal languages are not expressive enough. For instance, the natural language sentence

'There could be something that doesn't actually exist'

can only be rendered in a language comprising an **actuality operator A** as well as an existence predicate E [Hodes, 1984a; Hodes, 1984b], namely as $\Diamond\exists x \neg \mathbf{A} E(x)$.[22]

Extensions of the modal language, however, require modification to the Lewis translation scheme, as has been attempted for instance by Graeme Forbes [Forbes, 1982] and Murali Ramachandran [Ramachandran, 1989].

But these translations are only partially faithful, as Fara and Williamson argue [Fara and Williamson, 2005] who translate inconsistent sentences of modal predicate logic involving an actuality operator (interpreted in standard Kripke semantics) to satisfiable sentences of counterpart theory via different translation schemes. (Yet, this is not too surprising given that the semantics of counterpart theory is more flexible than the standard Kripkean semantics of modal predicate logic.)

---

[22] This example is related to an ontological objection raised by Plantinga [Plantinga, 1974] as well as Konyndyk [Konyndyk, 1986], namely that the standard interpretation of modal language implies the existence of fictional, nonexistent objects, which is at odds with several varieties of modal actualism. Compare also [Chihara, 1998].

We will not follow this line of research here any further. Rather, we are interested in the ideas underlying counterpart theory as a semantical framework for quantified modal logic in general, an approach that was initiated by Allen Hazen [Hazen, 1979]. From this perspective, counterpart theory can be reformulated as follows.

We assume that a frame is a collection of first-order structures over pairwise disjoint domains, and add a relation C on the union of the domains. If x and y are from the same domain, then $C(x, y)$ if and only if $x = y$. Thus, a **counterpart frame** is a pair $\langle W, C\rangle$, where W is a set of worlds such that

$$D_v \cap D_w = \varnothing \text{ iff } v \neq w, \tag{37}$$

and $C \subseteq (\bigcup_{w\in W} D_w)^2$ a relation such that, for each v

$$C \cap D_v^2 = \{\langle x, x\rangle : x \in D_v\}. \tag{38}$$

Instead of specifying actual objects we make use of the actual world. The formula $\Diamond\varphi(\vec{a})$ is true at w if there is a world v and counterparts c_i for a_i in v, i.e. $\langle a_i, c_i\rangle \in C(w, v)$ for all i, such that $\varphi(\vec{c})$ is true at v. The following are *not* theorems in counterpart frames, though they are valid in standard constant-domain Kripke-frames.

1. $(x = y) \to \Box(x = y)$
2. $(x \neq y) \to \Box(x \neq y)$
3. $\forall x\Box\varphi(x) \to \Box\forall x\varphi(x)$

The necessity of identity and distinctness fail because counterpart relations need not be functional nor do they need to be injective. Also, the Barcan formula is not valid for there could be individuals in an accessible world w that are not the counterpart of any individual in v (counterpart relations are not assumed to be 'surjective'). The converse Barcan formula $\Box\forall x\varphi(x) \to \forall x\Box\varphi(x)$, however, is also a theorem of counterpart theory, for if *everything* in w makes φ true so does a fortiori everything that is a counterpart of something in v.

Notice that Lewis considers only counterpart relation from one world to another world. However, there may be many 'different ways' to move from one world to another. This distinguishes Kripke-type semantics also significantly from standard counterpart theory (*cf.* [Lewis, 1968]) and its derived possible worlds semantics (*cf.* [Hazen, 1979]).

That this feature of multiple functions or relations is not eliminable is due to the fact that there are second-order closed MPLs that are complete only with respect to frames having at least two counterpart relations between worlds, *cf.* [Kracht and Kutz, 2005].

Counterparts need not be unique. An object can have several counterparts in another world. Neither are counterpart relations assumed in general to be symmetric or transitive. If counterparts were unique both ways, the theory would be exactly like that of Kripke's, only that we have changed the ontology. Objects in the Kripke-frames become equivalence classes of the counterpart relation. Although that seems to be just a technical move, Lewis did try to identify criteria for establishing whether an object is a counterpart of another. One principle may be explicated as follows (see [Forbes, 1985]).

> For $c \in D_u$ and $d \in D_v$, d is a counterpart of c only if nothing in v is more similar to c as it is in u than is d as it is in v.

This is only a necessary condition and it is not clear what would be sufficient. Furthermore, interpretations of counterparthood based on similarity face several problems in general. For instance, as Feldman discusses in [Feldman, 1971], how could we possibly translate a sentence such as 'I could have been quite unlike what I in fact am' when counterparthood is understood solely as a relation of similarity?

Another problem is that the counterparts are defined for each object individually. This creates technical problems. Suppose we only have equality, and axioms that say 'there are exactly three individuals' and 'there is exactly one world'. Then the counterpart relations must be functions, despite the fact that the above definition would predict that any object can serve as a counterpart for any given object. Another problem is this. If similarity is defined in terms of how many of the properties the objects share, we can only use 1-place properties to determine similarity and must exclude properties such as 'a is father of b'. However, once we have chosen a counterpart for b in the next world there is only one way to choose a counterpart for a, despite the fact that prior to the choice of b, we might have had more than one counterpart.

Thus, we see that we must actually specify the counterparts for all objects at once. This leads to the **presheaf semantics** outlined in [Ghilardi, 1991]. The original definitions worked only for extensions of **S4**, but they can be generalised easily. Consider a category $\mathcal{C}$. A category has objects, which are here considered as *worlds*, and **morphisms** or **maps** between objects, which define what we call **modes of transgression**. The idea is that from a world u there are several ways to transgress into the world v. Now, a presheaf is a functor F from $\mathcal{C}$ to the category of sets and functions. This means that every world u is mapped to a set $F(u)$, and every map $f : u \to v$ is mapped to a function $F(f) : F(u) \to F(v)$. There are two further conditions: $F(f \circ g) = F(f) \circ F(g)$, and for the identity id_u we have $F(id_u) = id_{F(u)}$. We consider for simplicity valuations at a single world. If u is a world, then a valuation β at u is a function from the object variables to $F(u)$. The clause for the modal operator is as follows.

$$\langle F, u, \beta\rangle \vDash \Diamond\varphi :\Leftrightarrow \text{ there is } f : u \to v \text{ such that } \langle F, v, F(f) \circ \beta\rangle \vDash \varphi \qquad (39)$$

Thus, one thinks of $F(f)$ as the function that takes each object in u into its counterpart in v under the mode of transgression f. If β is a valuation at u, and $f : u \to v$ a mode of transgression, then $F(f) : F(u) \to F(v)$ can be composed with β to give a valuation at v. The following is valid in presheaf semantics:

1. All axioms of **QS4** are tautologies.
2. $(x = y) \to \Box(x = y)$.
3. The tautologies are closed under (MP), Universal Generalisation and Leibniz' Principle.

[Kracht and Kutz, 2002] take one more step of generalisation. Rather than using functions, they employ once again relations, but allow a set of them. We give a slight adaptation of the definitions of that paper to take care of the fact that we now deal with possibilist quantifiers plus an existence predicate as opposed to 'proper' free-logical quantifiers. Furthermore, we shall make more explicit the world dependence of the universes.

Call a relation $R \subseteq M \times N$ a **CE-relation** (CE stands for 'counterpart existence') if for all $x \in M$ there exists a $y \in N$ such that $x \mathrel{R} y$ and, likewise, for all $y \in N$ there exists an $x \in M$ such that $x \mathrel{R} y$. Since we assume free logic, this is a harmless complication (it does not entail claims about existence, but is needed to avoid truth-value gaps).

Definition 4. A **poly-counterpart frame** is a quadruple $\langle W, T, \mathfrak{U}, \mathfrak{C}\rangle$, where $W, T \neq \varnothing$ are nonempty sets, $\mathfrak{U}$ a function assigning to each $v \in W$ a nonempty subset $\mathfrak{U}_v$ of T (its **domain**) and, finally, $\mathfrak{C}$ a function assigning to each pair of worlds v, w a set $\mathfrak{C}(v, w)$ of CE-relations from $\mathfrak{U}_v$ to $\mathfrak{U}_w$. A pair $\langle \mathfrak{W}, \mathfrak{I}\rangle$ is called a **counterpart structure** if $\mathfrak{W}$ is a counterpart frame and $\mathfrak{I}$ an interpretation, that is, a function assigning to each $w \in W$ and to each n-ary predicate letter a subset of $\mathfrak{U}_w^n$.

We say that v **sees** w in $\mathfrak{F}$ if $\mathfrak{C}(v, w) \neq \varnothing$—thus the notion of 'accessibility' is completely reduced to the existence of a counterpart relation. A **valuation** is a function η which assigns to every possible world v and every variable an element from the universe $\mathfrak{U}_v$ of v. We write η_v for the valuation η at v. A **counterpart model** is a quadruple $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{I}, \eta, w\rangle$, where $\mathfrak{F}$ is a counterpart frame, $\mathfrak{I}$ an interpretation, η a valuation and $w \in W$. If $\rho \in \mathfrak{C}(v, w)$, write $\eta \xrightarrow{\rho} \widetilde{\eta}$ if for all $x \in V$: $\langle \eta_v(x), \widetilde{\eta}_w(x)\rangle \in \rho$. Truth in a counterpart model is defined as follows. Let $v \in W$ and η be a valuation.

$$\begin{aligned}\langle \mathfrak{F}, \mathfrak{I}, \eta, v\rangle \vDash \Diamond\varphi(\vec{y}) :\Leftrightarrow\ & \text{there are } w \in W, \rho \in \mathfrak{C}(v, w) \text{ and } \widetilde{\eta} \\ & \text{such that } \eta \xrightarrow{\rho} \widetilde{\eta} \text{ and } \langle \mathfrak{F}, \mathfrak{I}, \widetilde{\eta}, w\rangle \vDash \varphi(\vec{y}) \qquad (40)\end{aligned}$$

This semantics can be enriched by towers in the sense above to obtain general structures. The minimal logic of these structures can be axiomatised as follows. On top of free logic and modal logic, we require the truth of the following principles for identity.

1. $x = x$.

2. If y is free for $\varphi(x)$, then $x = y \to (\varphi(x) \leftrightarrow \varphi(y//x))$, where $\varphi(y//x)$ denotes the result of replacing one or more free occurrences of x not in the scope of a modal operator by y.

3. $x = y \wedge \Diamond\top \to \Diamond(x = y)$.

The last postulate is true generally of counterpart theory. However, it can hardly be considered valid when one thinks of the quantifiers as quantifying over intensional rather than extensional (trace-like) objects (see the discussion of constants below). The necessity of identity, $(x = y) \to \Box(x = y)$, as well as the necessity of difference are both invalid. To see the first, let u be a world containing just a with two counterparts b and c in v. Then we can choose for x the counterpart b and for y the counterpart c.

Notice that for constants the situation is different. For suppose that the language contains constants and that they are interpreted by individual concepts. Then neither principle is valid. For let in the same model the interpretation of $\underline{d}$ be a in u and b in v, but for $\underline{e}$ it is a in u and c in v. Then $(\underline{d} = \underline{e}) \wedge \Diamond\top \to \Diamond(\underline{e} = \underline{e})$ as well as $(\underline{d} = \underline{e}) \to \Box(\underline{d} = \underline{e})$ are both invalid. This disparity of constants and variables is a rather worrying aspect of counterpart semantics. Presheaf semantics does not have this defect since counterparts are unique. On the other hand, in presheaf semantics the necessity of identity is valid.

The modal Leibniz' law of [Kracht and Kutz, 2002] allows for simultaneous substitution of all free occurrences of x by y in $\Diamond\chi$ (denoted by $\Diamond\chi(y//x)$, provided that $x = y$ is true.

$$\bigwedge x. \bigwedge y. x = y. \to .\Diamond\chi(x) \to \Diamond\chi(y//x) \tag{41}$$

Now, notice that in Kripke-semantics the rule of replacing constants for universally quantified variables is valid. In counterpart frames this creates unexpected difficulties.

For suppose that constants are present and may be substituted for variables. Then we may derive from (41), using the substitution of c for x and d for y:

$$(c = d \wedge \Diamond\top) \to \Diamond(c = d) \tag{42}$$

Since a constant has a fixed interpretation in each world, this means that if two constants are equal in a world and there exists some accessible world, then there will also be some accessible world in which they are equal. This is not generally valid. What is happening here is a shift from a *de re* to a *de dicto* interpretation. If we follow the traces of the objects, the formula is valid, but if we substitute intensional objects, namely constants, it becomes refutable. Notice that this situation is also reflected in the way non-rigid constants are treated in [Fitting and Mendelsohn, 1998]. There, the two possible readings of the above formula, the de dicto and de re reading, are distinguished by actually binding the interpretation of the constants to the respective worlds by using the term-binding λ-operator.[23]

Applied to Hesperus and Phosphorus, this means that if they are equal, then there is a belief world of George's in which they are equal. However, if George believes that they are different, this cannot be the case. So, the counterpart semantics cannot handle constants correctly—at least not in a straightforward way, i.e., without restricting the possible values of constants in accessible worlds. This paradox is avoided in Kracht and Kutz [2002] by assuming that the language actually has *no* constants.

Counterpart theory as well as modal predicate logic were originally intended to formalise natural language statements involving modalities, but their expressive richness make them also obvious candidates for knowledge representation in artificial intelligence. Thus, let us make a few comments about their computational properties.

Counterpart theory is formulated as a first-order theory. While the **monadic fragment** [Löwenheim, 1915] (containing only unary predicates) or the **two-variable fragment** [Scott, 1962; Mortimer, 1975] of classical predicate logic are decidable, the fragment of predicate logic with binary predicates and three variables is already undecidable [Surányi, 1943], which means that only a very limited fragment of counterpart theory is decidable.

When we move to modal predicate logic, though, the situation does not improve. Kripke already showed in 1962 that the monadic fragment of MPL is undecidable [Kripke, 1962]. Later, it was shown that also the two-variable fragment of practically all standard MPLs based on Kripkean constant domain semantics are undecidable [Gabbay and Shehtman, 1993], which seemed to imply that the decision problem for modal predicate logic is almost hopelessly difficult. Only fairly recently has the search for decidable fragments of MPL been successful, namely with the definition of so-called **monodic fragments**[24] [Wolter and Zakharyaschev, 2001]. These are languages of MPL where at most one free variable is allowed in the scope of a modal operator; and basically any extension of a decidable fragment of classical predicate calculus to its monodic modal language yields a decidable formalism [Gabbay *et al.*, 2003].

In general, more subtle combinations of languages have been studied, such as **fusions** [Kracht and Wolter, 1991; Fine and Schurz, 1991] or **products** [Gabbay *et al.*, 2003] of modal logics (where modal predicate logic can be understood as the product of first-order predicate logic and modal propositional logic). But again, while fusions behave very nicely both computationally and logically, they are rather inexpressive, and, on the other hand, products, being quite expressive, are once again computationally very difficult in general [Gabbay *et al.*, 2003].

---

[23] λ-abstraction was introduced to modal logic by Robert Stalnaker and Richmond H. Thomason in [Stalnaker and Thomason, 1968; Thomason and Stalnaker, 1968] and corresponds to Bertrand Russell's scoping device used to analyse **definite descriptions**, originating in [Russell, 1905] and systematically used in Principia Mathematica [Whitehead and Russell, 1910, 1912, 1913]. More recently, λ-abstraction was studied by Melvin Fitting [Fitting, 1999], who also investigated modal languages that comprise λ-abstraction for constants but without allowing quantification, thus defining languages that are situated between modal propositional and predicate logics [Fitting, 2002a]. For a discussion compare also [Braüner and Ghilardi, 2006].

[24] A 'monody' is a kind of music distinguished by having a single melodic line and accompaniment.

Interestingly, then, the general idea of counterpart relations being based on a notion of similarity also gives rise to a framework of knowledge representation languages that is rather expressive, natural from a semantical point of view, and which is very well-behaved computationally, namely the theory of **$\mathcal{E}$-connections** [Kutz, 2004; Kutz *et al.*, 2004]. In $\mathcal{E}$-connections, a finite number of formalisms talking about disjoint domains are 'connected' by relations relating entities in different domains, intended to capture different aspects or representations of the 'same object'. For instance, an 'abstract' object o of a description logic L_1 can be related via a relation R to its life-span in a temporal logic L_2 (a set of time points) as well as to its spatial extension in a spatial logic L_3 (a set of points in a topological space, for instance). As with poly-counterpart frames, the presence of multiple relations between domains is essential for the versatility of this framework, the expressiveness of which can be varied by allowing different language constructs to be applied to the connecting relations. $\mathcal{E}$-connections approximate the expressivity of products of logics 'from below' and could, perhaps, be considered a more 'cognitively adequate' version of counterpart theory.

8 INDIVIDUAL CONCEPTS

We have noticed above that in many analyses there is an asymmetry between variables and constants. This asymmetry is easily explained. A constant is standardly interpreted by an **individual concept**, while variables are often used to refer to objects. In counterpart theory, objects are world bound, so the notion of an object is different from that of an individual concept. There are proposals which argue that the values of variables should likewise be individual concepts. Recent proposals to this effect are the worldline frames of [Schurz, 1997] and the coherence frames of [Kracht and Kutz, 2005].

A **coherence frame** is a quintuple $\langle W, R, O, T, \tau\rangle$, where $\langle W, R, O\rangle$ is a predicate Kripke-frame, consisting of a set W of worlds, an accessibility relation R, and a set O, the set of **objects**, and where T is a set, the set of **things**, and $\tau : U \times W \to T$ a surjective function. We call τ the **trace function** and $\tau(o, w)$ the **trace of** o **in** w. The set $T_w = \{t \in T : t = \tau(o, w), o \in O\}$ is the set of things constituting the world w, i.e. the set of things which can bear properties. In [Schurz, 1997], a **worldline** is defined to be a function $W \to T$. It turns out that objects can be identified with worldlines, and we shall simplify our exposition accordingly.

An **interpretation** is a function $\mathfrak{I}$ mapping each predicate letter P of arity $\Omega(P)$ to a function from W to $T_w^{\Omega(P)}$ and each constant symbol c to a member of U. Let us call an interpretation $\mathfrak{I}$ **equivalential** if for all $\vec{a}, \vec{c} \in U^{\Omega(P)}$ and

$w \in W$, if $\tau(a_i, w) = \tau(c_i, w)$ for all $i < \Omega(P)$ then $\vec{a} \in \mathfrak{I}(P)(w)$ if and only if $\vec{c} \in \mathfrak{I}(P)(w)$. (This condition is enforced only for extensional predicates.) A **coherence structure**, then, is a pair $\langle \mathfrak{W}, \mathfrak{I} \rangle$ where $\mathfrak{W}$ is a coherence frame and $\mathfrak{I}$ an equivalential interpretation. A **coherence model** is a triple $\langle \mathfrak{C}, \beta, w \rangle$, where $\mathfrak{C}$ is a coherence structure, $\beta : V \to U$ a valuation, $w \in W$. Every term t evaluates into a unique object $\varepsilon(t)$ (which depends on the valuation).

$$\begin{array}{lll} \langle \mathfrak{C}, \beta, w \rangle \vDash P(\vec{t}) & :\Leftrightarrow & \varepsilon(\vec{t}) \in \mathfrak{I}_w(P) \\ \langle \mathfrak{C}, \beta, w \rangle \vDash s = t & :\Leftrightarrow & \tau(\varepsilon(s), w) = \tau(\varepsilon(t), w) \\ \langle \mathfrak{C}, \beta, w \rangle \vDash \bigvee x.\varphi & :\Leftrightarrow & \text{for some } \gamma \text{ with } \gamma \sim_x \beta : \langle \mathfrak{C}, \gamma, w \rangle \vDash \varphi \\ \langle \mathfrak{C}, \beta, w \rangle \vDash \Diamond\varphi & :\Leftrightarrow & \text{there is } w' \rhd w \text{ such that } \langle \mathfrak{C}, \beta, w' \rangle \vDash \varphi \end{array} \tag{43}$$

In worldline semantics, predicates are interpreted as sets of n-tuples of traces. The following is put in place of the first condition:

$$\langle \mathfrak{C}, \beta, w \rangle \vDash P(\vec{t}) :\Leftrightarrow \langle \tau(\varepsilon(t_1), w), \cdots, \tau(\varepsilon(t_n), w) \rangle \in \mathfrak{I}_w(P) \tag{44}$$

This allows to drop the condition of equivalentiality. However, it removes the flexibility in dealing with intensional predicates. It is a matter of straightforward verification to show that all axioms and rules of the minimal MPL are valid in a coherence frame. Moreover, the set of formulae valid in a coherence structure constitutes a first-order MPL. Notice that the fourth postulate for equality holds in virtue of the special clause for equality and the condition that the interpretation must be equivalential. For if $\langle \mathfrak{C}, \beta, w \rangle \vDash y_i = y_n$, then $\tau(\beta(y_i), w) = \tau(\beta(y_n), w)$. So, if $\langle \mathfrak{C}, \beta, w \rangle \vDash P(y_0, \ldots, y_{n-1})$ for $P \in \Pi$, then $\langle \beta(y_i) : i < n \rangle \in \mathfrak{I}(P)(w)$. Let $\beta' \sim_{y_i} \beta$ be such that $\beta'(y_i) = \beta(y_n)$. By equivalentiality, $\langle \beta'(y_i) : i < n \rangle \in \mathfrak{I}(P)(w)$. This means that $\langle \mathfrak{C}, \beta', w \rangle \vDash P(y_0, \ldots, y_{n-1})$, and so $\langle \mathfrak{C}, \beta, w \rangle \vDash [y_n/y_i]P(y_0, \ldots, y_{n-1})$. If $\mathfrak{F}$ is a coherence frame, put $\mathfrak{F} \vDash \varphi$ if $\langle \mathfrak{F}, \mathfrak{I} \rangle \vDash \varphi$ for all equivalential interpretations $\mathfrak{I}$. Evidently, $\{\varphi : \mathfrak{F} \vDash \varphi\}$ is a first-order MPL.

The frames allow two objects to share the entire worldline, just as worlds can support the same propositions. This means that in a coherence frame the objects are not mere constructions, they exist in their own right. We say that $o = o'$ is true at w if and only if $\tau(o)(w) = \tau(o')(w)$. It follows that the formula

$$x = y \wedge \Diamond\top \to \Diamond(x = y) \tag{45}$$

is not valid in coherence frames, since there is no way to predict how the worldline of an object develops from one world to the next. If we turn to predicates, their interpretation is now a function $\mathfrak{I}(P)$, which assigns to each world w a subset of D_w^n. Thus predicates, like identity, are predicated not of the objects themselves, but of the traces in their worlds. Thus if $\tau(o)(w) = \tau(o')(w)$ then o and o' satisfy the same nonmodal formulae. The semantics is generalised using towers. This leads to the notion of a **generalised coherence frame**. Free logic is assumed as well. [Kracht and Kutz, 2005] show that the semantics of generalised coherence frames is complete for all modal predicate logics. Moreover, similar to the case of standard Kripke semantics, a variant of the well-known Henkin construction yields a general completeness result with respect to structures involving interpretations (rather than with respect to frames). The difference, however, to standard Kripke semantics is that coherence frames also cover various logics of identity. For details compare [Kracht and Kutz, 2005].

This semantics eliminates on the one hand the asymmetry between variables and constants and on the other the possibility of speaking of things ('res') simpliciter. A thing can only be understood as the materialisation of an individual concept, and is not conceptualised itself. This is especially noticeable in the light of the criticism leveled against counterpart theory. It has namely been observed that counterpart theory makes every actual property of an object at a world w a necessary one, for the reason that no object of another world is identical to it (see [Plantinga, 1974] for the argument and [Norton, 1976] for a rebuttal). Thus, the condition that attributions are exclusively of the individual concept remove that problem. The semantics based on the interpretation of variables as individual concepts is complete and provides a rather elementary completeness proof for modal predicate logic. Also, the notion of an individual concept is not more suspicious than the counterpart theory. Counterpart relations can be seen as the traces of these functions. On the other hand, given a counterpart relation, we can define individual concepts in the following way. Given a counterpart relation C, a compatible worldline is any function f such that if $v \; R \; w$ then $f(v) \; C \; f(w)$. However, if $v \; R \; v$ it may become impossible to choose enough worldlines. However, one can always first unravel the counterpart frame before extracting the worldlines (see [Kracht and Kutz, 2005]).

Note that since trace functions are assumed to be surjective, every trace has to be the trace of some object. This is a natural condition, because objects are considered to be the primary entity, and traces a derived notion. The notion of equivalence is perhaps a curious one. It says that the basic properties of objects cannot discriminate between objects of equal trace. So, if Pierre believes that London is beautiful and Londres is not, while at the same time Londres *is* London [Kripke, 1979], we have two objects which happen to have the same trace in this world, though not in any of Pierre's belief worlds. Hence they must share all properties *in this world*. So, London and Londres can only be both beautiful or both ugly. This seems very plausible indeed. From a technical point of view, however, the fact that they cannot simply have different properties is a mere stipulation on our part. On the other hand, it is conceivable that there are basic predicates that are actually intensional, which would mean that they fail the substitution under (extensional) equality.

The difference with the counterpart semantics is that we have disentangled the quantification over objects from the quantification over worlds. Moreover, objects exist independently of worlds. Each object leaves a trace in a given world, though it need not exist there. Furthermore, two objects can have the same trace in any given world without being identical. However, identity of two objects holds in a world if and only if they have the same trace in it. If we also have function symbols, the clauses for basic predicates and equality will have to be generalised in the obvious direction.

9 OBJECTS IN COUNTERPART FRAMES

One of the main intuitions behind counterpart frames is that objects do not exist in more than one world, they are, in Lewis' terminology, **worldbound**. We call such objects **individuals**. Variables are interpreted in them as individuals, with no connection between the values at different worlds assumed. A more abstract, transworld notion of an **object** can only be reconstructed by following an individuals's counterparts along the counterpart relations being present. It turns out, however, that counterpart frames may have very few objects in this sense.

Definition 5 (Objects). Let $\mathfrak{F} = \langle W, T, \mathfrak{U}, \mathfrak{C} \rangle$ be a counterpart frame. An **object** is a function $f : W \to T$ such that (i) $f(v) \in \mathfrak{U}_v$ for all $v \in W$, (ii) for each pair $v, w \in W$ with $\mathfrak{C}(v, w) \neq \varnothing$ there is $\rho \in \mathfrak{C}(v, w)$ such that $\langle f(v), f(w) \rangle \in \rho$.

So, objects are constructed using the counterpart relation. If the trace b in world w is a counterpart of the trace a in world v, then there may be an object leaving trace a in v and trace b in w. If not, then not. However, there are frames which are not empty and possess no objects. Here is an example. Let $W = \{v\}$, $T = \{a, b\}$, $\mathfrak{U}_v = \{a, b\}$, and $\mathfrak{C}(v, v) = \{\rho\}$ with $\rho = \{\langle a, b \rangle, \langle b, a \rangle\}$. It is easy to see that this frame has no objects. The crux is that we can only choose one trace per world, but when we pass to an accessible world, we must choose a counterpart. This may become impossible the moment we have cycles in the frame.

Thus, the connection between coherence frames and counterpart frames is not at all straightforward. Since the logic of a counterpart frame is a first-order modal predicate logic, one might expect that for every counterpart frame there is a coherence frame having the same logic. This is only approximately the case. It follows from Theorem 7 below that for every counterpart *structure* there is a coherence structure having the same theory. This is not generally true for frames. However, adopting a modification of coherence frames proposed by Melvin Fitting in [Fitting, 2003], namely balanced coherence frames (in [Fitting, 2003] the corresponding frames are called **Riemann FOIL frames** in analogy to Riemann surfaces in complex analysis), it can indeed be shown that for every counterpart frame there is a balanced coherence frame validating the same logic (under a translation).

Let us begin by elucidating some of the connections between counterpart and coherence frames. Note again that since counterpart structures as defined above do not interpret constants, we have to assume that the language does not contain any constants.

First, fix a coherence structure $\mathfrak{C} = \langle W, \lhd, U, T, \tau, \mathfrak{I} \rangle$. We put $U_v := \{\tau(o, v) : o \in U\}$. This defines the domains of the world. Next, for $v, w \in W$ we put $\rho(v, w) := \{\langle \tau(o, v), \tau(o, w) \rangle : o \in U\}$ and $\mathfrak{C}(v, w) := \varnothing$ if $v \lhd w$ does not obtain; otherwise, $\mathfrak{C}(v, w) := \{\rho(v, w)\}$. Finally, $\langle \tau(a_i, w) : i < \Omega(P) \rangle \in \mathfrak{I}'(P)(w)$ iff $\langle a_i : i < \Omega(P) \rangle \in \mathfrak{I}(P)(w)$. Then $\langle W, T, \mathfrak{U}, \mathfrak{C}, \mathfrak{I}' \rangle$ is a counterpart structure. We shall denote it by $CP(\mathfrak{C})$. Notice that there is at most one counterpart relation between any two worlds.

Conversely, let a counterpart structure $\mathfrak{N} = \langle W, T, \mathcal{U}, \mathcal{C}, \mathfrak{I}\rangle$ be given. We put $v \lhd w$ iff $\mathcal{C}(v, w) \neq \varnothing$. $U := T$. Let O be the set of all objects $o : W \to T$. Further, $\tau(o, w) := o(w)$. This defines a coherence frame if the set of objects is nonempty.[25] Finally, $\langle o_i : i < \Omega(P)\rangle \in \mathfrak{I}'(P)(w)$ iff $\langle o_i(w) : i < n\rangle \in \mathfrak{I}(P)(w)$. It is easy to see that this is an equivalential interpretation. So, $\langle W, \lhd, O, U, \tau, \mathfrak{I}'\rangle$ is a coherence structure, which we denote by $CH(\mathfrak{N})$.

Unfortunately, the logical relation between these two types of structures is rather opaque, not the least since the notion of satisfaction in them is different. Moreover, the operations just defined are not inverses of each other. For example, as we have already seen, there exist counterpart structures with nonempty domains which have no objects. In this case $CP(CH(\mathfrak{N})) \not\cong \mathfrak{N}$. Also let $\mathfrak{C}$ be the following coherence frame. $W := \{v, w, x, y\}$, $T := \{1, 2, 3, 4, 5, 6\}$, $U = \{a, b\}$, $\lhd = \{\langle v, w\rangle, \langle w, x\rangle, \langle x, y\rangle\}$. Finally, $\tau(a, -) : v \mapsto 1, w \mapsto 2, x \mapsto 4, y \mapsto 5$, $\tau(b, -) : v \mapsto 1, w \mapsto 3, x \mapsto 4, y \mapsto 6$. Generating the counterpart frame we find that 2 and 3 are counterparts of 1, and 5 and 6 are counterparts of 4. Hence, there are more objects in the counterpart frame than existed in the coherence frame, for example the function $v \mapsto 1, w \mapsto 2, x \mapsto 4, y \mapsto 6$.

Definition 6 (Threads). Let $\mathfrak{N}$ be a counterpart frame. A sequence $\langle (w_i, t_i) : i < n\rangle$ is called a **thread** if (1) for all $i < n$: $w_i \in W$, $t_i \in \mathcal{U}_{w_i}$, and (2) for all $i < n - 1$: $w_i \lhd w_{i+1}$ and $\langle t_i, t_{i+1}\rangle \in \rho$ for some $\rho \in \mathcal{C}(w_i, w_{i+1})$. $\mathfrak{N}$ is **rich in objects** if for all threads there exists an object o such that $o(w_i) = t_i$ for all $i < n$.

Notice that if $\lhd$ has the property that any path between two worlds is unique then $\mathfrak{N}$ is automatically object rich. Otherwise, when there are two paths leading to the same world, we must be able to choose the same counterpart in it. Using unravelling one can always produce an object rich structure from a given one. Additionally, we can ensure that between any two worlds there is at most one counterpart relation. We call counterpart frames that satisfy the condition $|\mathcal{C}(v, w)| \leq 1$ for all worlds $v, w \in W$ **Lewisian counterpart frames**. The proof of the following theorem makes use of the unravelling technique.

Theorem 7. For every counterpart structure $\mathfrak{N}$ there exists a Lewisian counterpart structure $\mathfrak{N}'$ rich in objects such that $\mathfrak{N}$ and $\mathfrak{N}'$ have the same first-order modal theory.

For a proof, let $\mathfrak{N} = \langle\langle W, T, \mathcal{U}, \mathcal{C}\rangle, \mathfrak{I}\rangle$ be a counterpart structure. A **path** in $\mathfrak{N}$ is a sequence $\pi = \langle w_0, \langle\langle w_i, \rho_i\rangle : 0 < i < n\rangle\rangle$ such that $\rho_i \in \mathcal{C}(w_{i-1}, w_i)$ for all $0 < i < n$. We let $e(\pi) := w_{n-1}$ and $r(\pi) = \rho_{n-1}$ and call these, respectively, the **end point** and the **end relation** of π. Let W' be the set of all paths in $\mathfrak{N}$ and $T' := T$. Further, let $\mathcal{U}'_\pi := \mathcal{U}_{e(\pi)}$ and for two paths π and μ put $\mathcal{C}'(\pi, \mu) := r(\mu)$ if $r(\mu) \in \mathcal{C}(e(\pi), e(\mu))$ and empty otherwise. Finally, let P be an n-ary predicate letter. Then $\mathfrak{I}'(P)(\pi) := \mathfrak{I}(P)(e(\pi))$. Now let $\mathfrak{N}' = \langle\langle W', T', \mathcal{U}', \mathfrak{C}'\rangle, \mathfrak{I}'\rangle$. This is a Lewisian counterpart structure and clearly rich in objects. It can be verified by induction that if β is a valuation on $\mathfrak{N}$ and w a world, and if β' is a valuation on $\mathfrak{N}'$ and π a path such that $e(\pi) = w$ and $\beta'_\pi(x_i) = \beta_w(x_i)$, then $\langle \mathfrak{N}, \beta, w\rangle \vDash \varphi$ iff $\langle \mathfrak{N}', \beta', \pi\rangle \vDash \varphi$ for all φ.

---

[25] Strictly speaking, we have to reduce the set U of traces to those elements $t \in T$ that actually are the trace of some object o, but this makes no difference semantically.

Notice by the way that for second-order closed logics this theorem is false. This is so because the interpretation of a predicate in π must be identical to that of μ if the two have identical end points. We will re-encounter this problem in the next section when we will modify Kripkean semantics to match counterpart frames in terms of generality. However, if we are interested in characterising MPLs by means of models, it follows from the above result that we can restrict ourselves in the discussion to Lewisian counterpart structures that are rich in objects.

But we can also strike the following compromise. Let us keep the counterpart semantics as it is, but interpret formulae in a different way. We say that $\mathfrak{M} = \langle\langle \mathfrak{F}, \mathfrak{I}\rangle, \beta, v\rangle$ is an **objectual counterpart model**, if $\mathfrak{F}$ is a counterpart frame as before, $\mathfrak{I}$ is an **objectual interpretation**, that is, a counterpart interpretation that additionally assigns objects to constant symbols, β an **objectual valuation** into $\mathfrak{F}$, i.e., a function that assigns to each variable an object in a given world. In this context, $\varepsilon_v(o) := \beta_v(o)$ if o is a variable and $\varepsilon_v(o) = \mathfrak{I}(o)(v)$ if o is a constant symbol.

Write $\beta \to^{\vec{y}}_{v,w} \beta$ if for some $\rho \in \mathcal{C}(v, w)$ we have $\langle \beta_v(x_i), \beta_w(x_i)\rangle \in \rho$ for all $x_i \in \vec{y}$. Furthermore, write $\beta \to^{\vec{y}}_{v,w} \gamma$ if for some $\rho \in \mathcal{C}(v, w)$ we have $\langle \beta_v(x_i), \gamma_w(x_i)\rangle \in \rho$ for all $x_i \in \vec{y}$, where γ is an objectual valuation. Terms t_i denote either variables or constants, $\vec{y}$ tuples of variables and $\vec{c}$ tuples of constants. The symbol $\vDash^*$ is called the **weak objectual truth-relation** and is defined thus:

$$\begin{array}{lll}
\langle \mathfrak{M}, \beta, v\rangle \vDash^* R(\vec{t}) & :\Leftrightarrow & \langle \varepsilon_v(t_0), \ldots, \varepsilon_v(t_{n-1})\rangle \in \mathfrak{I}_v(R) \\
\langle \mathfrak{M}, \beta, v\rangle \vDash^* \Diamond\varphi(\vec{y}, \vec{c}) & :\Leftrightarrow & \text{there is } \beta \to^{\vec{y}}_{v,w} \gamma : \langle \mathfrak{M}, \gamma, w\rangle \vDash^* \varphi(\vec{y}, \vec{c}) \\
\langle \mathfrak{M}, \beta, v\rangle \vDash^* \bigvee y.\varphi(y, \vec{c}) & :\Leftrightarrow & \text{there is } \widetilde{\beta} \sim_y \beta : \langle \mathfrak{M}, \widetilde{\beta}, v\rangle \vDash^* \varphi(y, \vec{c})
\end{array} \tag{46}$$

The **strong objectual truth-relation** $\vDash^\dagger$ is like $\vDash^*$ except for the clause for $\Diamond$ which is now

$$\langle \mathfrak{M}, \beta, v\rangle \vDash^\dagger \Diamond\varphi(\vec{y}, \vec{c}) \quad :\Leftrightarrow \quad \text{there is } \beta \to^{\vec{y}}_{v,w} \beta : \langle \mathfrak{M}, \beta, w\rangle \vDash^\dagger \varphi(\vec{y}, \vec{c}) \tag{47}$$

These interpretations remove the asymmetry between variables and constants in the sense that constants and variables are now assigned the same kind of values. However, while the strong objectual interpretation brings us very close to coherence semantics, the weak interpretation still bears essential properties of counterpart semantics, namely that we may move via a counterpart relation to a new object. More precisely we have the following:

Proposition 8. The rule of substituting constants for universally quantified variables is valid in the strong objectual interpretation. More specifically, for every counterpart frame $\mathfrak{F}$

$$\mathfrak{F} \vDash^\dagger (\bigwedge x.\varphi) \to [c/x]\varphi. \tag{48}$$

Furthermore, there is an objectual counterpart model $\mathfrak{M}$ such that

$$\mathfrak{M} \not\models^{\dagger} \bigwedge x. \bigwedge y.(x = y) \rightarrow (\varphi(x, \vec{z}) \leftrightarrow \varphi(y, \vec{z})) \tag{49}$$

Both claims are false for the weak objectual interpretation.

In object rich frames, the weak objectual interpretation gives the same theorems as the standard interpretation:

Theorem 9. Let $\mathfrak{N}$ be a counterpart structure rich in objects, v a world and let β be an objectual valuation and $\widetilde{\beta}$ a counterpart valuation such that $\beta_v(x_i) = \widetilde{\beta}_v(x_i)$ for all variables. Then for all φ:

$$\langle \mathfrak{N}, \beta, v\rangle \models^{*} \varphi \quad \Leftrightarrow \quad \langle \mathfrak{N}, \widetilde{\beta}, v\rangle \models \varphi \tag{50}$$

10 DUAL ONTOLOGIES, OR, THE SEMANTICAL IMPACT OF HAECCEITISM

No two philosophers agree on the nature of individuals, even when restricted to the more mundane question of what a material object is. The same, unsurprisingly, holds true for the notion of a **modal individual**—a concept which some would even claim to be nonsensical or inconsistent, but which, nevertheless, is central to the metaphysics of modality.

For instance, the discussion concerning the question of how spatio-temporal objects can **persist** in time, currently divided between adherents of **endurantism** (also called **3-dimensionalism**, the view that an object exists at a time by being *wholly present* at that time, and *persists* by being wholly present at more than one time [Jackson, 1998]) versus **perdurantism** (also called **4-dimensionalism**, roughly, the view that a material object has temporal as well as spatial parts, and *persists* by having distinct temporal parts at different times), consists essentially in competing ontologies as concerns the notion of a spatio-temporal object [Parsons, 2000; Sider, 2001]—and these conceptions of object correspond roughly to the different ontologies as represented in coherence frames versus standard semantics.[26] Obviously, this discussion can be specialised to the question whether objects in Minkowski space-time should be considered as perduring or rather as enduring, where, prima facie, special relativity seems to support the four-dimensional ontology, compare [Balashov, 2000a; Balashov, 2000b].

Another metaphysical doctrine that has been subject to considerable inquiry is (first-order) **haecceitism**, the view that there might be worlds that are distinguishable only by what individuals play what roles,[27] first introduced by David Kaplan [Kaplan, 1975], and elaborated on for instance by Joseph Melia [Melia, 1999] in connection with determinism and the 'hole argument' of Earman and Norton [1987]. In [Melia, 1999], he defines **D-haecceitism** to be the view that a theory may be indeterministic, even if all the different possible futures open to any world which makes the theory true are qualitatively identical. In [Melia, 2001; Melia, 2003], Melia also discusses **second-order haecceitism**, which may be defined as the position that there could be distinct worlds that agree on which second-order properties are occupied, but which disagree on which properties play which roles.[28]

---

[26] There is some controversy on which terminology is more appropriate (or less confusing). David Lewis and Mark Johnstone used the endurantism/perdurantism distinction [Lewis, 1986; Johnston, 1987], while many later authors, including Theodore Sider [Sider, 2001], preferred the 3/4-dimensionalism distinction.

A recurrent theme of this article was the idea that a semantics for modal logic should make as little ontological commitments as possible. It should not be necessary to define a new semantics once you commit to a new theory of identity across worlds, for example. The most flexible semantics in this respect so far (without moving to an algebraic semantics) was the generalised counterpart theory embodied in the notion of a poly-counterpart frame. In the following, we analyse how we can achieve the same kind of semantical generality while sticking to the more conventional Kripkean picture of possible worlds semantics. Curiously, this modification of Kripkean semantics requires to build a notion of haecceitism into the frames, namely the notion of a **world-mirror**, see below.

By the theorems above we can introduce the notion of an object into counterpart frames, which then makes them rather similar to coherence frames. However, counterpart structures with object valuations are still different from coherence structures. A different approach is to translate $\Diamond$ in order to accommodate the truth relation $\vDash^*$ within the language of counterpart structures.

$$(\Diamond\varphi(y_0,\dots,y_{n-1}))^\gamma := \bigvee z_0.\dots\bigvee z_{n-1}.z_0 = y_0 \wedge z_1 = y_1 \wedge \dots \\ \dots \wedge z_{n-1} = y_{n-1} \wedge \Diamond\varphi(\vec{z}/\vec{y})^\gamma \quad (51)$$

Here, y_i $(i < n)$ are the free variables of φ and z_i $(i < n)$ distinct variables not occurring in φ. This is actually unique only up to renaming of bound variables. This translation makes explicit the fact that variables inside a $\Diamond$ are on a par with bound variables.

Proposition 10. Let $\mathfrak{N}$ be a counterpart structure and x a world. Then for any φ:

$$\langle\mathfrak{N}, x\rangle \vDash \varphi^\gamma \quad \Leftrightarrow \quad \langle\mathfrak{N}, x\rangle \vDash \varphi$$

In object rich structures also $\vDash$ and $\vDash^*$ coincide, which makes all four notions the same. So, while in counterpart structures the formulae φ and φ^γ are equivalent, they are certainly not equivalent when interpreted in coherence structures.

---

[27] As expected, philosophers disagree on this notion. According to [Melia, 2003], David Lewis [Lewis, 1986] and David Armstrong [Armstrong, 1989] should be regarded as anti-haecceitists, while the later Kaplan [Kaplan, 1975] is a haecceistist. Brian Skyrms, on the other hand, accepts haecceitism for actual entities, but not for merely possible ones [Skyrms, 1981].

[28] David Lewis and David Armstrong both accept *this* version of haecceitism.

In [Schurz, 1997], it is shown that worldline semantics provides for the same class of frame complete logics in the absence of extra equality axioms as standard constant domain semantics. It follows that the same holds for coherence frames. This means that while coherence frames allow for a more natural treatment of non-rigid designation for example, unlike counterpart frames, they do not enlarge the class of frame complete logics unless one moves to the full second-order semantics involving towers as sketched above. But there is a different approach to this problem. Instead of introducing algebras of admissible interpretations we can assume that certain worlds are isomorphic copies of each other. So, we add to the frames an equivalence relation between worlds and require that predicates are always interpreted in the same way in equivalent worlds. This idea is due to Melvin Fitting (see [Fitting, 2003]). Call a relation $\mathcal{E} \subseteq W \times W$ a **world-mirror on $\mathfrak{F}$** if $\mathcal{E}$ is an equivalence relation and whenever $v \ \mathcal{E} \ w$ and $v \lhd u_1$, there is a u_2 such that $w \lhd u_2$ and $u_1 \ \mathcal{E} \ u_2$. Intuitively, two mirrored worlds v and w may be understood as a situation seen from two different perspectives (because v and w may have 'different histories', but have the 'same future').

Definition 11. A pair $\langle \mathfrak{F}, \mathcal{E} \rangle$ is called a **balanced coherence frame**, if $\mathfrak{F} = \langle W, \lhd, U, T, \tau \rangle$ is a coherence frame and $\mathcal{E}$ is a world-mirror on $\mathfrak{F}$. An interpretation $\mathfrak{I}$ is called **balanced**, if it is equivalential and $\langle u_0, \ldots, u_{n-1} \rangle \in \mathfrak{I}_v(P)$ iff $\langle u_0, \ldots, u_{n-1} \rangle \in \mathfrak{I}_w(P)$ for all n-ary relations P and for all worlds v, w such that $v \, \mathcal{E} \, w$. A **balanced coherence model** is a triple $\langle \langle \mathfrak{B}, \mathfrak{I} \rangle, \beta, w \rangle$, where $\mathfrak{B}$ is a balanced coherence frame, $\mathfrak{I}$ a balanced interpretation, β a valuation and w a world.

Theorem 12. For every counterpart frame $\mathfrak{F}$ there exists a balanced coherence frame $\mathfrak{F}^*$ such that for all formulae φ:

$$\mathfrak{F} \vDash \varphi \quad \Leftrightarrow \quad \mathfrak{F}^* \vDash \varphi^\gamma. \tag{52}$$

The details of the proof are rather technical and can be found in [Kracht and Kutz, 2005]. This result has interesting consequences. For example, since counterpart semantics is frame complete with respect to all first-order extensions $\mathbf{Q}L$ of canonical propositional modal logics L (compare [Ghilardi, 1992]), the same holds true for the translation $\mathbf{Q}L^\gamma$ with respect to balanced coherence frames. Now, we noted above that coherence frames per se characterise the same logics as standard constant domain semantics if no extra equality axioms are involved. But it is known that already rather simple canonical propositional logics have frame incomplete predicate extensions. In [Cresswell, 2001] it is shown that to complete frame incomplete MPLs by adding appropriate axioms, one needs mixed de re formulae rather than substitution instances of purely propositional formulae. So, the above result gives a hint on where the source for frame incompleteness with respect to standard semantics is to be found. In particular, note that the translation $.^\gamma$ leaves propositional formulae untouched, whereas de re formulae of the form $\Diamond\varphi(y_0, \ldots, y_{n-1})$ are transformed into formulae $\bigvee z_0 . \ldots \bigvee z_{n-1} . \bigwedge_{i<n} z_i = y_i \wedge \Diamond\varphi(\vec{z}/\vec{y})^\gamma$, which are de re formulae involving

Table 2. Competing Ontologies

	poly-counterpart	balanced standard
accessibility	relations between individuals	relations between worlds
objects	worldbound individuals	global universe of objects
identity	between individuals	equivalence between objects
predication	locally	globally (but admissible)
haecceitism	no	yes, world-mirrors

equality.

So what we need if we want to use standard possible worlds semantics to characterise a large class of logics via frame completeness are basically three things: firstly, the distinction between trace and object, secondly a different understanding of the modal operator as given by $.^{\gamma}$, and, thirdly, the assumption that certain worlds are copies of each other.

Let us make this claim more explicit. Given a standard constant domain frame $\langle W, \lhd, U\rangle$, we may add, as before, an equivalence relation $\mathcal{E}$ relating worlds. Furthermore, we technically do not need traces but can add a family of equivalence relations $(\mu_w)_{w\in W}$ interpreting equality at each world. Let us call frames of the form $\mathfrak{F} = \langle W, \lhd, U, (\mu_w)_{w\in W}, \mathcal{E}\rangle$ **balanced standard frames**. An interpretation $\mathfrak{I}$ is called **admissible** if interpretations agree on worlds related by $\mathcal{E}$ and, moreover, they respect the equivalence relations μ_w in the sense that $\vec{a} \in \mathfrak{I}(w)(P)$ iff $\vec{b} \in \mathfrak{I}(w)(P)$ whenever $a_i \mu_w b_i$ for all i. We may think of objects being related by μ_w as indiscriminable with respect to world w and basic extensional predicates. Call a valuation $\widetilde{\gamma}$ a w**-ignorant** $\vec{x}$**-variant** of γ, if $\widetilde{\gamma}(x_i)\mu_w\gamma(x_i)$ for all $x_i \in \vec{x}$. The truth definition for balanced standard frames is as usual except for the equality and modal clauses, which are as follows:

- $\langle \mathfrak{F}, \mathfrak{I}, \gamma, w\rangle \vDash x = y$ iff $\gamma(x)\mu_w\gamma(y)$;
- $\langle \mathfrak{F}, \mathfrak{I}, \gamma, w\rangle \vDash \Diamond\varphi(\vec{x})$ iff there is a w-ignorant $\vec{x}$-variant $\widetilde{\gamma}$ and a world $v \rhd w$ such that $\langle \mathfrak{F}, \mathfrak{I}, \widetilde{\gamma}, v\rangle \vDash \varphi(\vec{x})$;

It should be rather clear that there is a bijective correspondence between balanced coherence frames and balanced interpretations on the one hand and balanced standard frames and admissible interpretations on the other. Furthermore, for every $\langle W, \lhd, U, T, \tau, \mathcal{E}\rangle$ there is a $\langle W, \lhd, U, (\mu_w)_{w\in W}, \mathcal{E}\rangle$ such that for all φ

$$\langle W, \lhd, U, T, \tau, \mathcal{E}\rangle \vDash \varphi^{\gamma} \text{ iff } \langle W, \lhd, U, (\mu_w)_{w\in W}, \mathcal{E}\rangle \vDash \varphi.$$

Hence, the following theorem is an immediate corollary to Theorem 12.

Theorem 13. For every poly-counterpart frame there is a balanced standard frame having the same logic.

Table 2 gives a comparison of the different ontologies embodied in poly-counterpart frames versus balanced standard frames. It shows that if we want to avoid counterpart theory but still strive for a semantics that is of the same generality, we arrive at a framework where we not only have to give up numerical identity and replace it with equivalence relations, but where we are also committed to a certain form of haecceitism, namely the one that corresponds to the notion of a world-mirror. (However, similar assumptions are arrived at in the many-worlds interpretation of quantum mechanics, although, obviously, with entirely different motivations, compare [Tegmark, 2004].)

11 METAFRAMES

The most radical shift is presented in the semantics proposed by Valentin Shehtman and Dmitrij Skvortsov in [Skvortsov and Shehtman, 1993], which we alluded to in §6.4. Once again, it turned out in retrospect that the idea had an algebraic precursor in the hyperdoctrines of [Lawvere, 1969]. Though hyperdoctrines were originally defined as model structures for **Int**, the generalisation to other logics is straightforward (see [Shirasu, 1998]). Consider the category Σ of natural numbers and mappings between them. A **modal hyperdoctrine** is a covariant functor H from Σ into the category MA of modal algebras. $H(n)$ may be thought of as the algebra of meanings of formulae containing n free variables. To be well–defined, H must satisfy among other the so–called **Beck-Chevalley-condition**, which ensures that cylindrification has the same meaning on all levels. We shall not spell this out in detail; instead, we shall look at the natural dual $F := (-)^+ \circ H$ of H, where $(-)^+$ is defined as in (14). This maps n into the frame dual of $H(n)$. F is a contravariant functor into the category GFr of generalised frames. The system $\{F(n) : n \in \mathbb{N}\}$ provides a 'tower' of admissible systems of sets of their respective frames. In general, a **metaframe** is simply a contravariant functor from the category Σ to the category of general frames. Here is a recipe to obtain a metaframe which is adequate for the logic L. For every natural number n, let $H(n)$ be the algebra of the formulae which contain only x_1 to x_n free (under logical equivalence); and let $H(\sigma)$ be the result of substituting via σ. This defines a hyperdoctrine whose logic is L. To obtain a metaframe whose logic is L, simply take $M := (-)^+ \circ H$. The world of $M(n)$ are the n-**types** of L, and they constitute the worlds. The maps $M(\sigma)^+$ are the pre-images under the substitutions. They are p-morphisms between these frames. In general, for a metaframe M, $M(0)$ may be thought of as the frame of worlds in the usual sense. The frame of 1-types is different. First, since there is a map $\sigma : 0 \to 1$ (the empty map), there is a p-morphism $M(\sigma) : M(1) \to M(0)$. This map associates a 0-type to every 1-type. We may view each member of $M(1)$ as a world bound individual, but only up to logical equivalence. Similarly, $M(2)$ contains the world bound pairs of individuals, up to equivalence. An interpretation is a function ξ which maps every n-ary predicate letter P to a member of the internal sets of $M(n)$. Let v be in $M(k)$ for some k, and let $\sigma : n \to k$.

$$\langle M, \xi, \beta, v\rangle \vDash P(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) :\Leftrightarrow \langle \beta(x_1), \ldots, \beta(x_k)\rangle \in \sigma(\xi(P)) \qquad (53)$$

Notice that worlds of $M(n)$ can support the truth or falsity of formulae only if the free variables of that formula are within $\{x_1, \ldots, x_n\}$. Thus, the clause for $\exists$ requires care, while the clause for $\Diamond$ is actually straightforward. The problem with the existential is that it requires to shift to another world v', since eliminating the quantifier makes a variable free that need not be supported by v.

The method yields a metaframe M for each logic L. Moreover, if the propositional counterpart of L is canonical, so is L (simply pass from $M(n)$ to the corresponding full frame). Thus, from a technical point of view this construction is rather well understood. Unfortunately, from a metaphysical point of view it is not. It dispenses completely with the notion of an object and in some sense also with that of an ordinary possible world. What we are left with is possible states-of-affairs relating n individuals. Notice that accessibility relations are defined over these states-of-affairs for every n. So, if a and b are such state-of-affairs at level n, then $a \mathrel{R} b$ means that b is possible for a. Now, a can be expanded to a member of $M(n+1)$, and it can be reduced to a member of $M(n-1)$, and these operations are p-morphisms. This allows to reconstruct some of the classical notions. First, we can identify worlds as members of $M(0)$. Given a world v and $n > 0$, the set $O_n(v) := M(n)^{-1}(v)$ is the set of n-tuples of v. Now, ideally we would like to think of $O_n(v)$ as the n-fold cartesian product of $O_1(v)$. If that is so, [Skvortsov and Shehtman, 1993] call the metaframe **cartesian**. [Bauer, 2000] has shown that under mild conditions it is possible to construct a cartesian metaframe with the same n-types for all n out of a given metaframe. It then follows that all modal predicate logics are complete with respect to cartesian metaframes.

12 ESSENCE AND IDENTITY

An object is said to have a property P **essentially** if whenever x fails to have P it also fails to exist.[29] It has P **accidentally** in a world w if it has P. [Plantinga, 1974] has criticised counterpart theory of eliminating the distinction between accidental and essential properties of an object. [Burgess, 1978] paints a different picture. In his view, we can certainly speak of sortal objects and essential properties that these objects have even if there is no identity in the strict sense. The problem is partly due also to the question of what an object actually is. In coherence frames objects are interpreted as individual concepts. Now, we effectively have two sorts of identification: two objects can be identical in this world iff they have the same trace, while transworld identity is simply identity as object (identity of the whole worldline). How this changes matter is best seen with the theory of sets. [Forbes, 1985], as well as [Fine, 1977], expose a theory of sets, which embodies the idea that the essence of the set X is the fact that it contains the elements that it contains (extensionality). Moreover, if the fact that it has the given objects as members is an essential property, then we should have the following postulate to start, called **membership rigidity**.

---

[29] A discussion of various notions of 'essence' can be found in [Fine, 1995].

$$x \in X \to \Box(x \in X), \qquad x \notin X \to \Box(x \notin X) \tag{54}$$

(Notice that since we do not work with a distinguished actual world, our principles need not be prefixed by a box.) Additionally, it is assumed that $x \in X$ implies the existence of both x and X (the **falsehood principle**):

$$(\forall x)(\forall X)(x \in X \to E(x) \wedge E(X)) \tag{55}$$

How can finally transworld identity be expressed? In counterpart theory we need to go back and forth between worlds, assuring us that if $x \in X$ in w, say, then this is the case in v (for the respective counterparts) and conversely. In coherence frames we only need to state the principle in its original form.

$$(\forall X)(\forall Y)(X = Y \leftrightarrow (\forall x)(x \in X \leftrightarrow x \in Y)) \tag{56}$$

However, in this reading identity depends on the world we are in. If X is the set containing the morning star and the evening star, and Y the set containing only the evening star, then $X = Y$ is true at w iff, in w, the trace of the evening star is identical to the trace of the morning star. Of course, it is possible to retroactively introduce a stronger notion of identity, as with objects (see [Scott, 1970] on identity).

Coherence frames and worldline frames alike use identity in trace as a criterion of identity at a world. However, from a metaphysical point of view the introduction of traces is a dubious thing. What for example, is the trace of Yuri Gagarin in this world right now? Or that of Clark Kent? How do we decide whether or not Clark Kent is Superman on the basis of traces? If one is troubled by traces, one may do without them. All it takes is to introduce a binary predicate Eq whose interpretation is an equivalence relation in every world. The interpretation of Eq will hardly be that of identity. This has caused great concern; however, it must be borne in mind that Eq is a predicate of objects, not of individuals. All it takes is to consider objects to exist independently of their realisation ('individual'). This move is virtually forced on us when we want to supply metaframe semantics for logics that do not validate the necessity of identity (see [Kracht and Kutz, 2005]).

13 ON THE STATUS OF THE MODAL LANGUAGE

[McLeod, 2001] exposes three views one can have with regards to modal language; one can be a **primitivist**, a **eliminativist** and a **reductionist**. A primitivist holds that talk about modality is not reducible to anything else, and that modal statements have the same status as nonmodal ones. They are meaningful and they do not lack truth values. A **reductionist** believes that modal talk can be eliminated by talk about something else, for example possible worlds. Such a reduction might proceed along the lines of the standard translation. Finally, an **eliminativist** denies the meaningfulness of modal talk altogether. For him, modal talk is meaningless. For example, one might hold that there is no way to attach verification procedures to modal statements in the way that one does with ordinary, nonmodal statements. From a formal point of view, the reductionist works with the presupposition that modal talk is just proxy for talk about possible worlds, using the standard translation (6). But we have seen that modal logic is incomplete; the second-order interpretation in Kripke-frames confuses certain logics (namely the incomplete logics with their completion). The reductionist thus is committed to the rejection of incomplete logics, unless he agrees to yet another entity, the internal sets. It is the task of the reductionist to either show that the need for these sets does not arise (by showing that the logic of the modal operator is complete), or, if he cannot do this, he is committed to a kind of entity that we have shown to be equivalent in character with that of the propositions themselves. This actually strengthens the primitivist position, since the primitivist can point out that these objects he had in the first place. We do not aim at discussing the eliminativist's position here. It is mostly rejected on the grounds that it fails to explain human communication. Suffice it to point out that even Quine did consider certain interpretations of modal operators as unproblematic, for example those relating to time and space. While he rejected an 'intensional ontology' proper, he accepted modal discourse about real-world objects [Quine, 1976]. In his view, this does not constitute transgression into another world. (This is implicit in other work as well.) It is unclear, though, which interpretation of modality requires the addition of possible worlds in this sense, and which one does not. For example, suppose future is indeterminate and you might either toss head or tail in the next moment. Now, if that moment arrives and you have tossed head, does the other outcome in which you tossed tail exist or not? And, if it exists, does it exist in another world or in this one? On the many worlds interpretation of quantum mechanics it is really part of this world (despite the name 'many worlds interpretation'). If so, most counterfactual reasoning is actually perfectly acceptable because it talks about how things are in some other history. If one is inclined to reject this line because it is in principle impossible to know what happens in another history, notice that **verificationism** frequently employs procedures that cannot be carried out even in principle. The formulation of these principles involves modal talk, as [McLeod, 2001] points out, so there seems to be something irreducible about modal language. Interestingly, notice that much of set theory is founded around the idea that every potentially infinite chain defines an actual object. The natural numbers exist as soon as we have the infinite ascending chain of numbers 1, 2, 3, $\cdots$. Georg Cantor addressed the distinction between potential and actual infinity by saying that the principle of potential infinity cannot be formulated without conceiving of the totality of numbers in the first place (for example, by saying that *for every number there is a larger number*).

Still, the interpretations of modality seem to present different degrees of ontological commitment. Also, they might operate on different lines. Epistemic modalities seem to be very different from what we may call 'objective modalities' in that the former may be at odds with Leibniz' Principle of substituting equivalent descriptions. Belief, for example, is highly problematic in this respect. Even if the axioms of the natural numbers imply that Fermat's conjecture is true, one may believe the first but not the second. On these grounds it seems that the attempt to model belief via possible worlds is highly problematic. Jaakko Hintikka [Hintikka, 1962] confronts the problem with the notion of *implicit belief*, but that is not helpful since it fails to analyse the notion we are after, namely that of belief, and substitutes it with a different notion.

In a similar vein, talk of objects and counterparts is not simply problematic as such. If we talk about time and space, the notion of an object *seems* to be clear even to those who reject the notion of modality. However, even if there is (seemingly or not) no talk of modality, the questions about the nature of objects do not disappear, they just take a different form. For example, why is this cat today the same as that cat yesterday? The answer might go like this. There is a continuous function from the location of that cat yesterday to the location of this cat today, and at every instance between these time points, this function yields the location of an object at that time point, and this object was a cat. (See [Quine, 1976] for a discussion.) However implausible it is to expect from anyone to really go by this definition (rather than just checking whether this cat looks or acts the same as that cat yesterday, supposing that our memory and perceptions are infallible etc.), the question is whether this answer always works. Consider, for example, artefacts, such as cars or bicycles. They can be the same even though some parts have been exchanged. It is not immediately clear how to apply the previous definition in the case where you give your bicycle to a mechanic and get it back tomorrow with some parts being exchanged. (Suppose your bicycle gets totally disassembled—you have to say that it disappeared at that moment.) If we accept that the bicycle you get back today is the same as the one you handed in yesterday, you are additionally in for a Sorites type argument (see [Forbes, 1985]). However this matter can be resolved, the answers to the questions of identity across time do not seem to be less challenging than those of identity across worlds in general [Simmons, 1991; Balashov, 2000b; Sider, 2001]. Additionally, the verificationist answer we gave before does not even work in principle. Notice that if quantum physics is correct, there is no way to check the continuity of the movement of any object, not just elementary particles (compare also [Da Costa and Krause, 1997]). There is, moreover, no object as such: there is just a vector in an infinite Hilbert space, of which the object is but one aspect (and the de Broglie wave another). Below the uncertainty threshold just about any monstrosity can happen (creating and destroying particles, for example). This means, at least in theory, that it is impossible in principle to track the exact movement of that cat yesterday until it becomes this cat today.

BIBLIOGRAPHY

[Alchourrón *et al.*, 1985] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change. *Journal of Symbolic Logic*, 50, 510–530, 1985.

[Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap. *Entailment: The Logic of Relevance and Necessity Vol 1.* Princeton University Press, 1975.

[Anderson *et al.*, 1992] A. R. Anderson, N. D. Belnap, and M. J. Dunn. *Entailment: The Logic of Relevance and Necessity Vol 2.* Princeton University Press, 1992.

[Aristotle, 1989] Aristotle. *Prior Analytics.* Hackett Publishing Company, Indianapolis/Cambridge, 1989. translated, with introduction, notes, and commentary, by Robin Smith.

[Armstrong, 1989] D. M. Armstrong. *A Combinatorial Theory of Possibility.* Cambridge University Press, Cambridge, 1989.

[Balashov, 2000a] Y. V. Balashov. Enduring and perduring objects in Minkowski space-time. *Philosophical Studies*, 99, 129–166, 2000.

[Balashov, 2000b] Y. V. Balashov. Persistence and space-time: philosophical lessons of the pole and barn. *The Monist*, 83(3), 321–340, 2000.

[Barcan, 1946] R. C. Barcan. A functional calculus of first order based on srict implication. *Journal of Symbolic Logic*, 11, 1–16, 1946.

[Barwise and Perry, 1983] J. Barwise and J. Perry. *Situations and Attitudes.* MIT Press, Cambridge (Massachussetts), 1983.

[Barwise and Perry, 1985] J. Barwise and J. Perry. Shifting situations and shaken attitudes. *Linguistics and Philosophy*, 8, 103–161, 1985.

[Bauer and Wansing, 2002] S. Bauer and H. Wansing. Consequence, counterparts and substitution. *The Monist*, 85(4), 483–497, 2002.

[Bauer, 2000] S. Bauer. Metaframes, Typen und Modelle der modalen Prädikatenlogik. Master's thesis, Fachbereich Mathematik, Humboldt–Universität zu Berlin, 2000.

[Becker, 1960] O. Becker. Zur Rekonstruktion des Kyrieuon Logos des Diodoros Kronos. In Josef Derbolav and Friedhelm Nicolin, editors, *Erkenntnis und Verantwortung: Festschrift für Theodor Litt zum 80. Geburtstag*, pages 250–263. Düsseldorf, 1960.

[van Benthem *et al.*, 2006] J. van Benthem, P. Blackburn, and F. Wolter, editors. *Handbook of Modal Logic.* Elsevier, Amsterdam, 2006.

[van Benthem, 1983] J. van Benthem. *Modal and Classical Logic.* Bibliopolis, 1983.

[van Benthem, 1986] J. van Benthem. *Essays in Logical Semantics.* D. Reidel Publishing Company, Dordrecht, 1986.

[Bigaj, 2001] T. Bigaj. Three-valued logic, indeterminism, and quantum mechanics. *Journal of Philosophical Logic*, 30, 97–119, 2001.

[Blackburn, 1993] P. Blackburn. Nominal tense logic. *Notre Dame Journal of Formal Logic*, 39, 56–83, 1993.

[Bocheński, 2002] J.. M. Bocheński. *Formale Logik.* Verlag Karl Alber, Freiburg, München, 2002. 5th edition, first edition 1956, english translation by Ivo Thomas, A History of Formal Logic, University of Notre Dame Press, 1961.

[Braüner and Ghilardi, 2006] T. Braüner and S. Ghilardi. First-order modal logic. In Johan van Benthem, Patrick Blackburn, and Frank Wolter, editors, *Handbook of Modal Logic.* Elsevier, Amsterdam, 2006.

[Brink and Rewitzki, 2002] C. Brink and I. Rewitzki. Three dual ontologies. *Journal of Philosophical Logic*, 31, 543–568, 2002.

[Brouwer, 1907] L. E. J. Brouwer. *Over de Grondslagen der Wiskunde.* PhD thesis, Amsterdam, 1907. Translation: "On the foundation of mathematics" in Brouwer, Collected Works, I, (A. Heyting ed.), 1975, North-Holland, Amsterdam, pp.11–101.

[Brouwer, 1908] L. E. J. Brouwer. De onbetrouwbaarheid der logische principes. *Tijdschrift voor Wijsbegeerte*, 2, 152–158, 1908. Translation: The unreliability of the logical principles, Ibid. pp. 107–111.

[Brown, 1988] M. A. Brown. On the logic of ability. *Journal of Philosophical Logic*, 17, 1–26, 1988.

[Bull, 1968] R. A. Bull. On possible worlds in propositional calculus. *Theoria*, 34, 171–182, 1968.

[Burgess, 1978] J. P. Burgess. The unreal future. *Theoria*, 44, 157–179, 1978.

[Burgess, 1979] J. P. Burgess. Logic and time. *Journal of Symbolic Logic*, 44, 566–582, 1979.

[Carnap, 1946] R. Carnap. Modalities and quantification. *Journal of Symbolic Logic*, 11, 33–64, 1946.

[Carnap, 1947] R. Carnap. *Meaning and Necessity. A Study in Semantics and Modal Logic.* The University of Chicago Press, Chicago, 1947.

[Chagrov and Zakharyaschev, 1997] A. Chagrov and M. Zakharyaschev. *Modal Logic.* Oxford University Press, Oxford, 1997.

[Chihara, 1998] C. Chihara. *The Worlds of Possibility. Modal Realism and the Semantics of Modal Logic.* Clarendon Press, Oxford, 1998.

[Church, 1951] A. Church. A formulation of the logic of sense and denotation. In P. Henle, H. M. Kallen, and S. K. Langer, editors, *Structure, Method and Meaning: Essays in Honor of H. M. Sheffer*, pages 3–24. The Liberal Arts Press, New York, 1951.

[Cocchiarella, 1975] N. B. Cocchiarella. On the primary and secondary semantics of logical necessity. *Journal of Philosophical Logic*, 4(1), 13–18, 1975.

[Cocchiarella, 2002] N. B. Cocchiarella. Logical necessity based on Carnap's criterion of adequacy. *Korean Journal of Logic*, 5(2), 1–21, 2002.

[Copeland, 2002] J. Copeland. The genesis of possible worlds semantics. *Journal of Philosophical Logic*, 31(2), 99–137, 2002.

[Cresswell, 2001] M. J. Cresswell. How to complete some modal predicate logics. In Michael Zakharyaschev et al., editor, *Advances in Modal Logic. Volume 2*, Stanford, 2001. CSLI Publications.

[Czelakowski, 2001] J. Czelakowski. *Protoalgebraic Logics.* Number 10 in Trends in Logic. Kluwer, 2001.

[Da Costa and Krause, 1997] N. C. A. Da Costa and D. Krause. An intensional Schrödinger logic. *Notre Dame Journal of Formal Logic*, 38(2), 179–195, 1997.

[Divers, 2002] J. Divers. *Possible Worlds.* The Problems of Philosophy. Routledge, London and New York, 2002.

[Dowty, 1977] D. Dowty. Towards a semantic analysis of verb aspect and the English *"imperfective progressive"*. *Linguistics and Philosophy*, 1, 54–78, 1977.

[Earman and Norton, 1987] J. Earman and J. Norton. What price space-time substantivalism? The hole story. *British Journal for the Philosophy of Science*, 38, 515–525, 1987.

[Everett III, 1956] H. Everett III. *The theory of the universal wave function.* PhD thesis, Princeton University, 1956.

[Fara and Williamson, 2005] M. Fara and T. Williamson. Counterparts and actuality. *Mind*, 114, 1–30, 2005.

[Feldman, 1971] F. Feldman. Counterparts. *The Journal of Philosophy*, 68, 406–409, 1971.

[Fine and Schurz, 1991] K. Fine and G. Schurz. Transfer theorems for stratified modal logics. In *Proceedings of the Arthur Prior Memorial Conference*, Christchurch, New Zealand, 1991.

[Fine, 1975] K. Fine. Some connections between elementary and modal logic. In Stig Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 15–31. North–Holland, 1975.

[Fine, 1977] K. Fine. Properties, propositions, and sets. *Journal of Philosophical Logic*, 6, 135–191, 1977.

[Fine, 1981] K. Fine. Model theory for modal logic. Part III: existence and predication. *Journal of Philosophical Logic*, 10, 293–307, 1981.

[Fine, 1995] K. Fine. Senses of essence. In [Sinnott-Armstrong, 1995, pp. 53–73].

[Fitting and Mendelsohn, 1998] M. Fitting and R. L. Mendelsohn. *First-Order Modal Logic.* Kluwer Academic Publishers, Dordrecht, 1998.

[Fitting, 1999] M. Fitting. On quantified modal logic. *Fundamenta Informaticae*, 39(1–2), 105–121, 1999.

[Fitting, 2002a] M. Fitting. Modal logics between propositional and first-order. *Journal of Logic and Computation*, 12, 1017–1026, 2002.

[Fitting, 2002b] M. Fitting. *Types, Tableaus and Gödel's God*, volume 12 of *Trends in Logic.* Kluwer Academic Publishers, Dordrecht, 2002.

[Fitting, 2003] M. Fitting. First-order intensional logic. In D. Niwinski and J. Wolenski, editors, *Proccedings of the Alfred Tarski Centenary Conference.* Journal of Pure and Applied Logic (Special Issue), 2003.

[Forbes, 1982] G. Forbes. Canonical counterpart theory. *Analysis*, 42, 33–37, 1982.

[Forbes, 1985] G. Forbes. *The Metaphysics of Modality.* Oxford Clarendon Press, Oxford, 1985.

[van Fraassen, 1974] B. C. van Fraassen. Hidden variables in conditional logic. *Theoria*, 40, 176–190, 1974.

[Gabbay, 1972] D. M. Gabbay. A general theory of the conditional in terms of a ternary operator. *Theoria*, 38, 97–104, 1972.

[Gabbay and Shehtman, 1993] D. Gabbay and V. Shehtman. Undecidability of modal and intermediate first-order logics with two individual variables. *Journal of Symbolic Logic*, 58, 800–823, 1993.

[Gabbay *et al.*, 2003] D. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyaschev. *Many-Dimensional Modal Logics: Theory and Applications.* Number 148 in Studies in Logic and the Foundations of Mathematics. Elsevier Science Publishers, Amsterdam, 2003.

[Gallin, 1975] D. Gallin. *Intensional and Higher-Order Modal Logic.* North-Holland, Amsterdam, 1975.

[Garson, 1991] J. W. Garson. Applications of Free Logic to Quantified Intensional Logic. In Karel Lambert, editor, *Philosophical Applications of Free Logic*, pages 111–144. Oxford University Press, New York, Oxford, 1991.

[Gerson, 1975a] M. Gerson. An extension of S4 complete for neighbourhood semantics but incomplete for relational semantics. *Studia Logica*, 34, 333–342, 1975.

[Gerson, 1975b] M. Gerson. The inadequacy of neighborhood semantics for modal logics. *Journal of Symbolic Logic*, 40(2), 141–147, 1975.

[Ghilardi, 1991] S. Ghilardi. Presheaf semantics and independence results for some non-classical first-order logics. *Archiv für mathematische Logik*, 29, 125–136, 1991.

[Ghilardi, 1992] S. Ghilardi. Quantified extensions of canonical propositional intermediate logics. *Studia Logica*, 51, 195–214, 1992.

[Girard, 1987] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50, 1–102, 1987.

[Gödel, 1932] K. Gödel. Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien*, 69, 65–66, 1932.

[Gödel, 1933] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4, 39–40, 1933.

[Gödel, 1995] K. Gödel. Ontological proof. In Soloman Feferman, Jr. John W. Dawson, Warren Goldfarb, Charles Parsons, and Robert N. Solovay, editors, *Kurt Gödel Collected Works, Volume III, Unpublished Essays and Lectures*, pages 403–404. Oxford University Press, New York, 1995.

[Goldblatt, 1980] R. I. Goldblatt. Diodorean modality in Minkowski spacetime. *Studia Logica*, 39, 219–236, 1980.

[Goldblatt, 1992] R. I. Goldblatt. *Logics of Time and Computation.* Studies in Logic, Language and Information. CSLI, Stanford, 1992. second edition.

[Goldblatt, 2003] R. I. Goldblatt. Mathematical modal logic: A view of its evolution. *Journal of Applied Logic*, 1, 309–392, 2003.

[Hazen, 1976] A. Hazen. Expressive completeness in modal languages. *Journal of Philosophcal Logic*, 5, 25–46, 1976.

[Hazen, 1979] A. Hazen. Counterpart-theoretic semantics for modal logic. *The Journal of Philosophy*, 76, 319–338, 1979.

[Henkin, 1950] L. Henkin. Completeness in the Theory of Types. *Journal of Symbolic Logic*, 15, 81–91, 1950.

[Heyting, 1930] A. Heyting. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, pages 42–56, 1930.

[Hintikka, 1962] J. Hintikka. *Knowledge and Belief. An Introduction into the logic of the two notions.* Cornell University Press, Ithaca, 1962.

[Hintikka, 1964] J. Hintikka. Aristotle and the 'Master Argument' of Diodorus. *American Philosophical Quarterly*, 1, 101–114, 1964.

[Hintikka, 1969] J. Hintikka. Existential presuppositions and their elimination. In *Models for Modalities. Selected Essays*, pages 23–44. D. Reidel Publishing Company, Dordrecht, 1969.

[Hodes, 1984a] H. Hodes. Axioms for actuality. *Journal of Philosophical Logic*, 13, 27–34, 1984.

[Hodes, 1984b] H. Hodes. Some theorems on the expressive limitations of modal languages. *Journal of Philosophical Logic*, 13, 13–26, 1984.

[Hodes, 1986] H. Hodes. Individual actualism and three valued modal logics. Part I: model-theoretic semantics. *Journal of Philosophical Logic*, 15, 369–401, 1986.

[Holton, 2003] R. Holton. David Lewis's philosophy of language. *Mind & Language*, 18(3), 286–295, 2003.

[Hughes and Cresswell, 1996] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, London, 1996.

[Humberstone, 1981] L. Humberstone. From worlds to possibilities. *Journal of Philosophical Logic*, 10, 313–339, 1981.

[Jackson, 1998] F. Jackson. Metaphysics by possible cases. In *Mind, Method, and Conditionals: Selected Essays*, pages 133–153. Routledge, 1998.

[Johnston, 1987] M. Johnston. Is there a problem about persistence? *Aristotelean Society Supp.*, 61, 107–135, 1987.

[Jónsson and Tarski, 1951] B. Jónsson and A. Tarski. Boolean algebras with operators. *American Journal of Mathematics*, 73, 891–939, 1951.

[Kaplan, 1975] D. Kaplan. How to Russell a Frege-Church. *Journal of Philosophy*, 72, 716–729, 1975. reprinted in Loux, 1979.

[Kneale and Kneale, 1984] W. Kneale and M. Kneale. *The Development of Logic*. Clarendon Press, Oxford, 1984. paperback edition, first edition 1962.

[Konyndyk, 1986] K. Konyndyk. *Introduction to Modal Logic*. University of Notre Dame Press, Notre Dame, Indiana, 1986.

[Kracht and Kutz, 2002] M. Kracht and O. Kutz. The semantics of modal predicate logic I. counterpart-frames. In Frank Wolter, Heinrich Wansing, Maarten de Rijke, and Michael Zakharyaschev, editors, *Advances in Modal Logic. Volume 3*. World Scientific Publishing Company, 2002.

[Kracht and Kutz, 2005] M. Kracht and O. Kutz. The semantics of modal predicate logic II. Modal individuals revisited. In Reinhard Kahle, editor, *Intensionality*, volume 22 of *ASL Lecture Notes in Logic*, pages 60–97. A. K. Peters, Los Angeles, 2005.

[Kracht and Wolter, 1991] M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56, 1469–1485, 1991.

[Kracht and Wolter, 1999] M. Kracht and F. Wolter. Normal monomodal logics can simulate all others. *Journal of Symbolic Logic*, 64, 99–138, 1999.

[Kracht, 1999] M. Kracht. *Tools and Techniques in Modal Logic*. Number 142 in Studies in Logic. Elsevier, Amsterdam, 1999.

[Kracht, 2004] M. Kracht. Notes on substitution in first-order logic. In Uwe Scheffler, editor, *First-Order Logic 75*, pages 155–172, Berlin, 2004. Logos Verlag.

[Kracht, 2006] M. Kracht. Modal consequence relations. In Johan van Benthem, Patrick Blackburn, and Frank Wolter, editors, *Handbook of Modal Logic*. Elsevier, Amsterdam, 2006.

[Kratzer, 1989] A. Kratzer. An investigation into the lumps of thought. *Linguistics and Philosophy*, 12, 607–653, 1989.

[Kripke, 1959] S. A. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24(1), 1–14, 1959.

[Kripke, 1962] S. A. Kripke. The undecidability of monadic modal quantificational theory. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 8, 113–116, 1962.

[Kripke, 1963] S. A. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16, 83–94, 1963.

[Kripke, 1972] S. A. Kripke. Naming and necessity. In Gilbert Harman and Donald Davidson, editors, *Semantics of Natural Language*, number 40 in Synthese Library, pages 243–355. Reidel, Dordrecht, 1972.

[Kripke, 1979] S. A. Kripke. A puzzle about belief. In A. Margalit, editor, *Meaning and Use*, pages 239–283. D. Reidel Publishing Co., Dordrecht, Boston, 1979.

[Kutz *et al.*, 2003] O. Kutz, H. Sturm, N.-Y. Suzuki, F. Wolter, and M. Zakharyaschev. Logics of metric spaces. *ACM Transactions on Computational Logic (TOCL)*, 4(2), 260–294, 2003.

[Kutz *et al.*, 2004] O. Kutz, C. Lutz, F. Wolter, and M. Zakharyaschev. $\mathcal{E}$-Connections of abstract description systems. *Artificial Intelligence*, 156(1), 1–73, 2004.

[Kutz, 2000] O. Kutz. Kripke–Typ Semantiken für die modale Prädikatenlogik. Master's thesis, Fachbereich Mathematik, Humboldt–Universität zu Berlin, 2000.

[Kutz, 2003] O. Kutz. New semantics for modal predicate logics. In Benedikt Löwe, Wolfgang Malzkorn, and Thoralf Räsch, editors, *Foundations of the Formal Sciences II: Applications of Mathematical Logic in Philosophy and Linguistics*, volume 17 of *Trends in Logic*, pages 151–162, Dordrecht, Boston, London, 2003. Kluwer Academic Publishers.

[Kutz, 2004] O. Kutz. *$\mathcal{E}$-Connections and Logics of Distance*. PhD thesis, The University of Liverpool, 2004.

[Lambert, 1963] K. Lambert. Existential import revisited. *Notre Dame Journal of Formal Logic*, 4, 288–292, 1963.

[Lambert, 1997] K. Lambert. *Free Logics: Their Foundations, Character, and Some Applications Thereof.* Academia Verlag, Sankt Augustin, 1997.

[Lawvere, 1969] F. W. Lawvere. Adjointness in foundations. *Dialectica*, 23(3/4), 281–296, 1969.

[Lewis and Langford, 1932] C. I. Lewis and C. H. Langford. *Symbolic Logic.* Appleton-Century-Crofts, New York, 1932.

[Lewis, 1918] C. I. Lewis. *A Survey of Symbolic Logic.* University of California Press, Berkeley, 1918.

[Lewis, 1920] C. I. Lewis. Strict implication—an emendation. *Journal of Philosophy, Psychology, and Scientific Methods*, 17, 300–302, 1920.

[Lewis, 1968] D. K. Lewis. Counterpart theory and quantified modal logic. In *Journal of Philosophy, Vol. 65*, pages 113–126. repr. in *The Possible and the Actual*, ed. Michael J. Loux, Ithaca 1979; also in: David K. Lewis, *Philosophical papers 1*, Oxford 1983, 1968.

[Lewis, 1971] D. K. Lewis. Completeness and decidability of three logics of counterfactual conditionals. *Theoria*, 36, 74–85, 1971.

[Lewis, 1973] D. K. Lewis. *Counterfactuals.* Blackwell, Oxford, 1973.

[Lewis, 1986] D. K. Lewis. *On the Plurality of Worlds.* Blackwell, 1986.

[Leyton, 1992] M. Leyton. *Symmetry, Causality, Mind.* MIT Press, Cambridge (Mass.), 1992.

[Lindström and Segerberg, 2006] S. Lindström and K. Segerberg. Modal logic and philosophy. In Johan van Benthem, Patrick Blackburn, and Frank Wolter, editors, *Handbook of Modal Logic.* Elsevier, Amsterdam, 2006.

[Linsky, 1971] L. Linsky, editor. *Reference and Modality.* Oxford University Press, Oxford, 1971.

[Loux, 1979] M. J. Loux, editor. *The Possible and the Actual: Readings in the Metaphysics of Modality.* Cornell University Press, Ithaca and London, 1979.

[Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, 76, 447–470, 1915.

[Łukasiewicz, 1957] J. Łukasiewicz. *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic.* Oxford University Press, Oxford, 1957. 2nd edition.

[Marcus *et al.*, 1962] R. Barcan Marcus, W. V. Quine, S. Kripke, D. Føllesdal, and J. McCarthy. Discussion of the paper of Ruth B. Marcus. *Synthese*, XIV, 132–143, 1962. reprinted in [Marcus, 1993].

[Marcus, 1961] R. Barcan Marcus. Modalities and intensional languages. *Synthese*, XIII(4), 303–322, 1961. reprinted in [Barcan, 1993].

[Marcus, 1993] R. Barcan Marcus. *Modalities: Philosophical Essays.* Oxford University Press, Oxford, 1993.

[McLeod, 2001] S. McLeod. *Modality and Anti-Metaphysics.* Ashgate, Aldershot, 2001.

[Melia, 1999] J. Melia. Holes, haecceitism and two conceptions of determinism. *British Journal for the Philosophy of Science*, 50, 639–664, 1999.

[Melia, 2001] J. Melia. Reducing possibilities to language. *Analysis*, 61(1), 19–29, 2001.

[Melia, 2003] J. Melia. *Modality.* Central Problems of Philosophy. Acumen, 2003.

[Montague, 1968] R. Montague. Pragmatics. In B. Klibansky, editor, *Contemporary Philosophy: A Survey*, pages 102–122. La Nuova Italia Editrice, 1968.

[Montague, 1969] R. Montague. On the nature of certain philosophical entities. *Monist*, 53, 159–194, 1969. reprinted in [Montague, 1974; Chapter 5].

[Montague, 1974] R. Montague. *Formal Philosophy: Selected Papers of Richard Montague*, edited by Richmond H. Thomason. Yale University Press, New Haven and London, 1974.

[Mortimer, 1975] M. Mortimer. On languages with two variables. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 21, 135–140, 1975.

[Nelson, 1949] N. D. Nelson. Constructible falsity. *The Journal of Symbolic Logic*, 14(1), 16–26, 1949.

[Norton, 1976] B. Norton. Is counterpart theory inadequate? *Journal of Philosophical Logic*, 5, 79–89, 1976.

[Ohlbach *et al.*, 2001] H. J. Ohlbach, A. Nonnengart, M. de Rijke, and D. M. Gabbay. Encoding two-valued nonclassical logics in classical logic. In John Alan Robinson and Andrei Voronkov, editors, *Handbook of Automated Reasoning*, pages 1403–1486. Elsevier and MIT Press, 2001.

[Orlov, 1928] I. E. Orlov. The calculus of compatibility of propositions. *Mathematics of the USSR, Sbornik*, 35, 263–286, 1928. (Russian).

[Parry, 1970] W. T. Parry. In memoriam: Clarence Irving Lewis (1883–1964). *Notre Dame Journal of Formal Logic*, 11(2), 129–140, 1970.

[Parsons, 1995] T. Parsons. Ruth Barcan Marcus and the Barcan formula. In [Sinnott-Armstrong, 1995, pp. 3–11].

[Parsons, 1980] T. Parsons. *Nonexistent Objects.* Yale University Press, New Haven and London, 1980.

[Parsons, 2000] J. Parsons. Must a four-dimensionalist believe in temporal parts? *The Monist*, 83(3), 399–418, 2000.

[Patterson, 1995] R. Patterson. *Aristotle's Modal Logic. Essence and Entailment in the Organon.* Cambridge University Press, Cambridge, 1995.

[Pearce and Wagner, 1990] D. Pearce and G. Wagner. Reasoning with negative information I: strong negation in logic programs. In *Language, Knowledge and Intentionality (Acta Philosophica Fennica 49)*, pages 405–419. Helsinki, 1990.

[Pearce and Wagner, 1991] D. Pearce and G. Wagner. Logic programming with strong negation. In Peter Schroeder-Heister, editor, *Extensions of Logic Programming*, number 475 in Lecture Notes in Artificial Intelligence, pages 311–326. Springer, 1991.

[Perry, 1986] J. Perry. From worlds to situations. *Journal of Philosophical Logic*, 15, 83–107, 1986.

[Plantinga, 1974] A. Plantinga. *The Nature of Necessity.* Oxford Clarendon Press, 1974.

[Prior, 1955] A. N. Prior. Diodoran modalities. *Philosophical Quarterly*, 5, 205–212, 1955.

[Prior, 1957] A. N. Prior. *Time and Modality.* Oxford University Press, Oxford, 1957.

[Prior, 1967] A. N. Prior. *Past, Present and Future.* Oxford University Press, Oxford, 1967.

[Quine, 1947] W. V. O. Quine. The problem of interpreting modal logic. *Journal of Symbolic Logic*, 12(2), 43–48, 1947.

[Quine, 1953] W. V. O. Quine. Three grades of modal involvement. In *Proceedings of the XIth International Congress of Philosophy.* North Holland, Amsterdam, 1953. Reprinted in *The Ways of Paradox*, Random House, New York, 1966.

[Quine, 1976] W. V. O. Quine. Worlds away. *Journal of Philosophy*, 73, 859–863, 1976.

[Ramachandran, 1989] M. Ramachandran. An alternative translation scheme for counterpart theory. *Analysis*, 49, 131–141, 1989.

[Reynolds, 2002] M. Reynolds. Axioms for branching time. *Journal of Logic and Computation*, 12, 679–697, 2002.

[Reynolds, 2003] M. Reynolds. An axiomatization of Prior's Ockhamist logic of historical necessity. In Philippe Balbiani, Nobu-Yuki Suzuki, Frank Wolter, and Michael Zakharyaschev, editors, *Advances in Modal Logic, Vol. 4*, pages 355–370. King's College Publications, London, 2003.

[Rosen, 1990] G. Rosen. Modal fictionalism. *Mind*, 99, 327–354, 1990.

[Russell, 1905] B. Russell. On denoting. *Mind*, 14, 479–493, 1905.

[Schmidt, 2002] K. J. Schmidt. *Die modale Syllogistik des Aristoteles.* Mentis Verlag, Paderborn, 2002.

[Schurz, 1997] G. Schurz. *The Is-Ought Problem: An Investigation in Philosophical Logic*, volume 1 of *Trends in Logic.* Kluwer Academic Publishers, Dordrecht, 1997.

[Schurz, 1998] G. Schurz. Tarski and Carnap on logical truth—or: what is genuine logic. In Jan Wolenski and Eckehart Köhler, editors, *Alfred Tarski and the Vienna Circle—Austro-Polish Connections in Logical Empiricism*, pages 77–94. Kluwer Academic Publishers, Dordrecht, 1998.

[Schurz, 2001] G. Schurz. Carnap's modal logic. In W. Stelzner and M. Stöckler, editors, *Zwischen traditioneller und moderner Logik. Nichtklassische Ansätze*, pages 365–380. Mentis Verlag, Paderborn, 2001.

[Scott, 1962] D. Scott. A decision method for validity of sentences in two variables. *Journal of Symbolic Logic*, 27, 477, 1962.

[Scott, 1970] D. Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic. Some Recent Developments*, pages 143–174. Reidel, Dordrecht, 1970.

[Segerberg, 1971] K. Segerberg. An essay in classical modal logic. Mimeograph, Uppsala, 1971.

[Sendlewski, 1984] A. Sendlewski. Some investigations of varieties of N–lattices. *Studia Logica*, 43, 257–280, 1984.

[Sendlewski, 1990] A. Sendlewski. Nelson algebras through Heyting ones. *Studia Logica*, 49, 106–126, 1990.

[Shapirovsky and Shehtman, 2003] I. Shapirovsky and V. Shehtman. Chronological future modality in Minkowski spacetime. In Philippe Balbiani, Nobu-Yuki Suzuki, Frank Wolter, and Michael Zakharyaschev, editors, *Advances in Modal Logic, Vol. 4*, pages 437–459. King's College Publications, London, 2003.

[Shirasu, 1998] H. Shirasu. Duality in superintuitionistic and modal predicate logics. In Marcus Kracht et. al., editor, *Advances in Modal Logic, Volume 1*, pages 223–236. CSLI Publications, Stanford, 1998.

[Sider, 2001] T. Sider. *Four-Dimensionalism: An Ontology of Persistence and Time.* Clarendon Press, Oxford, 2001.

[Sider, 2002] T. Sider. The Ersatz pluriverse. *Journal of Philosophy*, 99, 279–315, 2002.

[Simmons, 1991] P. M. Simmons. On being spread out in time: temporal parts and the problem of change. In Wolfgang Spohn et. al., editor, *Existence and Explanation*, Dordrecht, 1991. Kluwer Academic Publishers.

[Sinnott-Armstrong, 1995] W. Sinnott-Armstrong, ed. *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus.* Cambridge University Press, Cambridge, 1995.

[Skvortsov and Shehtman, 1993] D. P. Skvortsov and V. B. Shehtman. Maximal Kripke–type semantics for modal and superintuitionistic predicate logics. *Annals of Pure and Applied Logic*, 63, 69–101, 1993.

[Skyrms, 1981] B. Skyrms. Tractarian nominalism. *Philosophical Studies*, 40, 199–206, 1981.

[Smith, 1998] Q. Smith. Direct, rigid designation and a posteriori necessity: a history and critique . In James Fetzer and Paul Humphreys, editors, *The New Theory of Reference: Kripke, Marcus and its Origins*, pages 137–178. Kluwer Academic Publishers, Dordrecht, 1998.

[Stalnaker and Thomason, 1968] R. Stalnaker and R. H. Thomason. Abstraction in first-order modal logic. *Theoria*, 34, 203–207, 1968.

[Stalnaker and Thomason, 1970] R. Stalnaker and R. H. Thomason. A semantic analysis of conditional logics. *Theoria*, 36, 23–42, 1970.

[Stalnaker, 1986] R. Stalnaker. Possible worlds and situations. *Journal of Philosophical Logic*, 15, 109–123, 1986.

[Surányi, 1943] J. Surányi. Zur Reduktion des Entscheidungsproblems des logischen Funktionenkalküls. *Mathematikai és Fizikai Lapok*, 50, 51–74, 1943.

[Suzuki, 1997] N.-Y. Suzuki. Kripke frames with graded accessibility and fuzzy possible world semantics. *Studia Logica*, 59, 249–269, 1997.

[Tarski, 1936] A. Tarski. Über den Begriff der logischen Folgerung. *Actes du Congrés International de Philosophie Scientifique*, 7, 1–11, 1936. translated as 'On the Concept of Logical Consequence' in [Tarski, 1956; pp. 409–420].

[Tarski, 1956] A. Tarski. *Logics, Semantics, Metamathematics*, Clarendon Press, Oxford, 1956.

[Tegmark, 2004] M. Tegmark. Parallel universes. In John D. Barrow, Paul C. W. Davies, and Charles L. Harper, editors, *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity*, pages 459–491. Cambridge University Press, Cambridge, 2004.

[Thom, 1979] P. Thom. Aristotle's syllogistic. *Notre Dame Journal of Formal Logic*, 20(4), 751–759, 1979.

[Thom, 1996] P. Thom. *The Logic of Essentialism: An Interpretation of Aristotle's Modal Syllogistic.* Kluwer, Dordrecht, 1996.

[Thomason and Stalnaker, 1968] R. H. Thomason and R. Stalnaker. Modality and reference. *Noûs*, 2, 359–372, 1968.

[Thomason, 1972] S. K. Thomason. Semantic analysis of tense logic. *Journal of Symbolic Logic*, 37, 150–158, 1972.

[Thomason, 1973] S. K. Thomason. A new representation of S5. *Notre Dame Journal of Formal Logic*, 14(2), 281–284, 1973.

[Troelstra, 1992] A. S. Troelstra. *Lectures on Linear Logic.* CSLI, Stanford, 1992.

[Vakarelov, 1977] D. Vakarelov. Notes on N–lattices and constructive logic with strong negation. *Studia Logica*, 36, 109–125, 1977.

[von Weizsäcker, 1985] C. F. von Weizsäcker. *Der Aufbau der Physik.* Carl Hanser Verlag, München, 1985.

[von Weizsäcker, 1992] C. F. von Weizsäcker. *Zeit und Wissen.* Carl Hanser Verlag, München, 1992.

[White, 1984] M. J. White. The necessity of the past and modal-tense logic incompleteness. *Notre Dame Journal of Formal Logic*, 25(1), 59–71, 1984.

[Whitehead and Russell, 1910, 1912, 1913] A. N. Whitehead and B. Russell. *Principia Mathematica.* Cambridge University Press, Cambridge, 1910, 1912, 1913. 3 volumes.

[Wójcicki, 1988] R. Wójcicki. *Theory of Logical Calculi.* Kluwer, Dordrecht, 1988.

[Wolter and Zakharyaschev, 2001] F. Wolter and M. Zakharyaschev. Decidable fragments of first-order modal logics. *Journal of Symbolic Logic*, 66, 1415–1438, 2001.

[Zanardo, 1985] A. Zanardo. A finite axiomatization of the set of strongly valid Ockhamist formulas. *Journal of Philosophical Logic*, 14, 447–468, 1985.

MODAL REALISM AND ITS ROOTS IN MATHEMATICAL REALISM

Charles S. Chihara

INTRODUCTION

Modal Realism is the doctrine that, besides the actual world we live in, there are other possible worlds. Someone who believes that such possible worlds exist is called a Modal Realist. Since the thesis that there exist other possible worlds is something that most non-philosophers find utterly fantastic, one might well wonder what would convince prominent philosophers to become Modal Realists. One consideration that has led some philosophers to adopt such a metaphysical position is to be found in Twentieth Century developments in modal logic. Briefly (and roughly), belief in possible worlds became attractive to a number of prominent philosophers because reference to such worlds figured so prominently in the mathematical semantics of modal logic — a kind of semantics that appeared to be both logically and philosophically fruitful.[1]

In what follows, I shall explain how certain plausible theses about the nature of mathematics were used to support key doctrines of two of our most eminent modal realists: David Lewis and Alvin Plantinga. I shall argue, in particular, that one central idea — the idea that the theorems of mathematics are genuine truths about the world — underlies key parts of the above Modal Realists' view of modality.

In this paper, I shall not only explain how such central doctrines about mathematics support the above philosophers' modal views, but I shall also argue that, if these supporting doctrines can be undermined, then much of the attractiveness of these versions of modal realism will dissolve. Then, I shall provide reasons for rejecting the doctrines about mathematics that underlie the metaphysical views about modality of these two philosophers.

In his *An Enquiry Concerning Human Understanding*, David Hume raised a powerful and influential objection to the view that metaphysicians are able to obtain *a priori* knowledge of matters of fact. Hume's reasoning is based upon a general principle which he argued had no exceptions. This is the principle that asserts that no "matter of fact" can be known *a priori*. For the empiricist Hume,

---

[1] For a more detailed discussion of this idea, see [Chihara, 1998, 1–3].

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

"matters of fact" can only be known on the basis of sense experience, so if a proposition is known *a priori*, he held that the proposition would have to be what he called a "relation of ideas". Using contemporary terminology, we can express Hume's position in roughly the following way: *synthetic* propositions can only be known on the basis of sense experience, so if some proposition is truly known a priori, then it would have to be an *analytic* truth.[2]

According to Hume's analysis, the only *analytic* truths that can be known at all are those of mathematics and logic. Thus, Hume concluded his *An Enquiry Concerning Human Understanding* with the following words:

> When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.[3]

Awakened from his dogmatic slumbers by passages such as the above, Kant responded by arguing, in his *Prolegomena to Any Future Metaphysics,* that the truths of mathematics are *synthetic*, not *analytic* as Hume had thought. Since Hume allowed that we can have *a priori* knowledge of the truths of mathematics, Kant could argue that Hume was simply mistaken in his belief that synthetic propositions can only be known on the basis of sense experience. If Hume had seen that mathematics yields *a priori* knowledge of synthetic truths, Kant suggests, the "good company into which metaphysics would thus have been brought would have saved it from the danger of a contemptuous ill-treatment, for the thrust intended for it must have reached mathematics, which was not and could not have been Hume's intention."[4]

1 DAVID LEWIS'S POSSIBLE WORLDS

In his book *On the Plurality of Worlds*, David Lewis responds to a serious objection that had been raised to his modal realism in a way that is strikingly similar to Kant's response to Hume. To explain Lewis's response to the objection, a brief description of his metaphysical views on modality will now be provided.

Lewis accepts the widely accepted logical analysis of the modal operators in terms of possible worlds according to which:

---

[2] Hume did not state his argument in terms of the expressions 'analytic' and 'synthetic', but rather in terms of 'relations of ideas' and 'matters of fact'. However, since Kant expressed his response to Hume using the more contemporary terminology, I have expressed Hume's objection using the Kantian terminology.

[3] David Hume, *An Enquiry Concernng Human Understanding* (La Salle: Open /court, 1949): p. 184.

[4] Immanuel Kant, *Prolegomena to Any Future Metaphysics* (New York: Liberal Arts Press, 1951): p. 18.

it is possible that ϕ
is true iff
there is a possible world in which ϕ is true.

What is striking about Lewis's analysis, however, is his understanding of the consequent of the above biconditional. The clause should not be understood to be a figurative or metaphorical way of explaining the operators. According to Lewis, the clause expresses a straightforward existential condition: it says that there literally exists a world in which that proposition is true — a world, perhaps, very much like the one we live in, even containing planets on which real flesh and blood animals live and die.[5] It follows from such an analysis of modal statements that, if the modal statements we assume to be true are indeed true, then our "world" — the vast universe we live in — is just one of many such "possible worlds" that exist.

This account of possible worlds gives rise to the following problem: Supposedly, there are these different possible worlds, each one spatiotemporally isolated from all the others; but presumably only one of these is the actual world. What distinguishes the actual world from the other merely possible ones? If all these possible worlds exist, what makes one of them the actual world? Or to put it another way: What does the actual world possess that the merely possible worlds lack?[6] It cannot be existence, since all the possible worlds are supposed to exist.

Lewis has a simple response. What 'actual' means depends upon who the speaker is: If the speaker is **a**, then 'the actual world' refers to the one and only world in which **a** exists. This doctrine is called the 'indexical analysis of actuality', since 'actual' is regarded as an indexical term, in the way 'I', 'this', 'here', and 'now' are regarded as indexicals: What these terms refer to depends upon features of the context of utterance [Lewis, 1986, 92]. In the case of 'actual', the feature of the context of utterance that is relevant is the world in which the utterance is made. From God's point of view — assuming with Leibniz that God is not an inhabitant of any world — all of the possible worlds are on a par. There is no special world that is singled out as *actual.* Of course, from our point of view, our world is special. It is actual. But that is analogous to saying that, from my point of view, *here* is special: Here is where I am.

Possible objects are worldbound

I noted above that, according to Lewis, a speaker exists in one and only one world. This may seem paradoxical. After all, according to standard versions of the semantics of modal logic, a person may exist in many possible worlds. But in the semantical analysis developed by Lewis, each person is "world-bound": If person **a** exists in world w, and if world $u \neq w$, then **a** does not exist in u. From the perspective of Modal Realism, such a doctrine is surely right: If all these possible worlds exist, and I am in one of them, how could I possibly be in any other world? It is easy enough to see how there could be someone very much like me in some other world; but no matter how much like me this person might be, he could not *be* me if he is in another world. That just could not be me, since I am in this world and he is in that world.

---

[5] Thus, Lewis writes: "When I profess realism about possible worlds, I mean to be taken literally. Possible worlds are what they are, and not some other thing. If asked what sort of thing they are, I cannot give any kind of reply my questioner probably expects: that is, a proposal to reduce possible worlds to something else.

I can only ask him to admit that he knows what sort of thing our actual world is, and then explain that other worlds are more things of *that* sort, differing not in kind but only in what goes on at them." [Lewis, 1979, 183–4].

[6] For a general discussion of this problem, see [Adams, 1979, 190–209].

Counterpart theory

The sentence 'Al Gore could have won the presidential election' is true, according to standard versions of possible worlds semantics, if and only if there is a possible world at which Al Gore does win the presidential election. Since the above sentence is clearly true, there must be a possible world, according to standard versions of possible worlds semantics, in which Gore exists and does things. How, then, does Lewis allow for the truth of such sentences, given his doctrine that every person (indeed, every concrete thing) is world-bound? According to Lewis's account of modality, the above sentence is true if and only if there is a possible world in which there is a *counterpart* of Gore who wins the presidential election.

What is a counterpart of Gore? A counterpart of Gore in world w resembles Gore — indeed, resembles Gore more closely than does anything else that exists in w — but it is not Gore. But how does this counterpart of Gore represent Gore as doing anything. Well, military personnel can use tiny models of soldiers, tanks, trucks, and artillery to represent a battle. Similarly, a possible world can have a real flesh and blood person to represent Gore: "[B]y waving his arm, or by winning the presidential election, the other-worldly [Gore] represents the this-worldly [Gore] as waving or as winning" [Lewis, 1986, 194].[7]

Not surprisingly, Lewis's views have met with a number of objections. In the following section, I shall take up one of the most common of these objections.

The unknowability objection[8]

According to Lewis's analysis, in order to know if any contrary-to-fact possibility statement is true, we would need to know that there literally exists a possible world of the sort characterized by Lewis. But such universes are completely beyond our sensory grasp, since they are causally isolated from us: none of our scientific instruments, no matter how advanced, could ever detect such a thing. Thus, the objector argues, all such modal statements could never be known by us to be true. And that surely is an absurdity.[9]

---

[7] For additional details of Lewis's counterpart theory, see [Chihara, 1998, 78–81].

[8] The objection discussed in this section is called the "Unknowability Objection" in [Chihara, 1998, 87], where it is presented in much more detail.

[9] See [Chihara, 1998, 87-9], for references to works raising this objection.

Lewis's response to the objection brings to mind Kant's response to Hume which was discussed earlier: he puts forward mathematical knowledge as a case in which "we can know that there exist countless objects causally isolated from us and unavailable to our inspection" [Lewis, 1986, 109]. Having thus convinced himself that we can have knowledge of causally inaccessible things, Lewis sees no insuperable barrier to obtaining knowledge of the existence and contents of other causally inaccessible things such as possible worlds. Indeed, one might interpret Lewis to be reasoning as follows. The reason supporters of the Unknowability Objection think we could not have any knowledge of other worlds is because they believe that it is not possible to have any knowledge of causally inaccessible things. But mathematical knowledge of sets refutes this belief: Sets are causally inaccessible, and yet we have knowledge of sets. By thus putting knowledge of possible worlds in the good company of mathematical knowledge, he thought he could save his own analysis of modality from the contemptuous ill-treatment implicit in the Unknowability Objection, writing: "Causal accounts of knowledge are all very well in their place, but if they are put forward as *general* theories, then mathematics refutes them".[10]

Lewis's mathematical realism

It can be inferred from the above discussion that Lewis was both a Modal Realist (since he espoused belief in the existence of possible worlds) and also a Mathematical Realist (since he believed in the existence of mathematical objects). It can also be seen that Lewis's Modal Realism rests in a way upon his Mathematical Realism, since his refutation of a crucial objection to his modal realism presupposes the truth of his Mathematical Realism.

Notice that, in the case of the two responses considered above, *theorems of mathematics are understood to express genuine truths.* Both Kant and Lewis thought that the theorems of mathematics assert what is the case, that is, they express straightforward truths. Kant believed that mathematical theorems assert synthetic truths, just as the scientific assertions of physicists assert synthetic truths. Lewis thought that mathematical theorems state truths about mathematical objects — entities that are causally isolated from us and unavailable to our inspection.

Let us divide Lewis's belief that mathematical theorems assert truths about mathematical objects into two sub-beliefs: (1) mathematical theorems are true assertions; and (2) mathematical theorems are assertions about mathematical objects. It is clear that Lewis believed both (1) and (2) to be true. Basically, that is why he believed that nominalists would have to maintain that mathematics is false. Nominalists do not believe in the existence of mathematical objects such as sets or classes. Thus, Lewis believes, the nominalist is committed to the position that the axioms and theorems of mathematics are false, thereby implying that mathematics should be rejected. So Lewis regards the nominalist as rejecting mathematics for philosophical reasons — something that Lewis considered an absurdity:

---

[10] [Lewis, 1986, 109]. It should be noted that Lewis's discussion of the unknowability objection involved several other points not discussed above.

> Rejecting classes means rejecting mathematics. That will not do. Mathematics is an established, going concern. Philosophy is as shaky as can be. To reject mathematics on philosophical grounds would be absurd [Lewis, 1991, 58].

Let us now consider why Lewis believed both (1) and (2). Start with (2), assuming, for now, that Lewis has strong reasons for believing (1). Why did Lewis think that mathematical theorems are assertions about mathematical objects? Evidently, it was because Lewis took the sentences of mathematics at face value. Under the most straightforward reading of the sentences of, say, set theory, the theorems of set theory are statements about sets. Furthermore, the standard logical formalizations of set theory yield just such a view of its theorems.

Consider now (1). Why did Lewis believe that the theorems of set theory are true statements? I do not believe that Lewis ever seriously considered the possibility that any genuine theorems of a well-established mathematical theory could fail to be true. (He once wrote: "How would you like to go and tell the mathematicians that they must change their ways, and abjure countless errors, now that philosophy has discovered that there are no classes? . . . Not me!"[11]) In fact, few philosophers have seriously questioned the view that mathematical theorems are true. For example, John Burgess and Gideon Rosen signal their acceptance of the central assumption in [Burgess and Rosen, 1997] when they advance a version of a Realistic view of mathematics which they call "minimal anti-nominalism"(a view they clearly advocate). "Having studied Euclid's Theorem," they write, "we are prepared to say that there exist infinitely many prime numbers". Furthermore, they understand Euclid's theorem in the standard straightforward way that logicians tend to understand such theorems, writing: "Moreover, when we say that there exist infinitely many prime numbers, we say so without conscious mental reservations or purpose of evasion. . ." [Burgess and Rosen, 1997, 10–1]. Thus, they are willing to accept the implications of their acceptance of the truth of the theorem by also asserting that there are prime numbers (and hence that mathematical objects exist).

In view of the fact that so many philosophers accept and believe, without any serious investigation, the thesis under consideration — the thesis that the theorems of mathematics are true — I shall call this thesis "the central assumption of philosophy of mathematics" or " the central assumption" for short.

Why the central assumption leads to mathematical realism

How does the central assumption lead to the Mathematical Realist's belief in the existence of mathematical objects? Using the Burgess-Rosen reasoning above as a model, we can develop an explicit argument for the existence of mathematical objects. In the following discussion, I shall use the term ‘classical mathematics’ to include set theory, number theory, and the theory of functions of real and complex numbers).

---

[11] [Lewis, 1991, 58].

Argument T:
(Central assumption) The theorems of classical mathematics are true.
(E) Some of the theorems of classical mathematics assert the existence of mathematical objects.
(Conclusion) There are mathematical objects.

Reasons for believing the central assumption

Why should we accept the central assumption? If one wants reasons for accepting it, I shall give, in the following discussion, some of the principal ones.[12] Lewis once claimed: “Modal Realism is fruitful; that gives us good reason to believe that it is true” [Lewis, 1986, 4]. Thus, it would be natural for him also to claim that, since the axioms and theorems of classical mathematics are extremely fruitful, that would give us good reason to believe that they are true. Certainly classical mathematics is constantly being fruitfully used by scientists in ways that suggests to us that the assertions of mathematics are truths. In particular, mathematics is *applied* in science and engineering to draw conclusions upon which we rely in our dealings with the world and other people. If the mathematical theorems were not true, how could we rationally place so much confidence in the conclusions that we infer using mathematics? Are not the theorems of analysis used time and time again in drawing empirically verifiable conclusions from scientifically established statements. If the mathematical theorems used were not true, would we not be basing our inferences upon falsehoods and thus shouldn’t we expect the bridges we build using mathematics to collapse, shouldn’t the rockets we program using mathematics go off course, etc.? Perhaps one can see why Michael Resnik would include the following as a premise in an argument he once produced:

> (R-1) *We are justified in drawing conclusions*
> *from and within science only if we are justified in taking the*
> *mathematics used in science to be true.*[13]

Resnik did not attempt to supply a convincing justification for accepting (R-1), evidently because he believed that the claim is so obviously true.[14]

Another related reason for accepting the central assumption is based upon the fact that many mathematical theorems seem to be directly verified to be true. Consider, for example, the *Fundamental Theorem of the Integral Calculus.*

---

[12] Other reasons are given in [Chihara, 2004], e.g. see “Maddy’s mystery” (pp. 229-30, 291-2).

[13] [Resnik, 1997, 48].

[14] Actually, Resnik does put forward a kind of argument for (R-1) in [Resnik, 1998, 233], but there he uses the premise that “science assumes . . . the truth of much mathematics”, which is itself not justified.

For any real-valued function of a real variable, $f(x)$,
which is continuous in the interval (a, b),
if $F(x)$ is an indefinite integral of $f(x)$, then
$\int_a^b f(x)dx = F(b) - F(a)$

This theorem not only has been successfully applied in countless ways for hundreds of years, evidently it has also been confirmed in a variety of empirical ways by students and professionals alike, by graphing specific continuous functions and empirically determining (or estimating) the areas under the graphs, say on engineering paper. For example, it would seem that what the theorem says, regarding the definite integral of the linear function $f(x) = 2x + 7$, with $a = 3$ and $b = 7$, can be checked, empirically, by graphing the function in the interval (3, 7) and counting the relevant squares.[15]

A closely related reason for accepting the central assumption is the undeniable fact that mathematical theorems frequently convey valuable information to scientists, engineers and researchers of many types (as the example of the Fundamental Theorem illustrates) — a fact which further suggests that such theorems are true, since declarative sentences that convey valuable information are, it would seem, for the most part true.

Lewis's version of the indispensability argument

Interestingly, even though Lewis believed the central assumption and certainly accepted (E), Lewis did not put forward a version of Argument T to justify his belief in the existence of sets. Instead, he attempted to justify his belief in sets by giving a "cost-benefit" version of Quine's Indispensability Argument. More specifically, he argued that belief in the existence of sets is justified by the theoretical benefits that accrue to one's overall theory of the world by assuming their existence. The following quotation expresses Lewis's reasoning:

> We have only to believe in the vast hierarchy of sets, and there we find entities suited to meet the needs of all the branches of mathematics; and we find that the very meagre primitive vocabulary of set theory, definitionally extended, suffices to meet our needs for mathematical predicates; and we find that the meagre axioms of set theory are first principles enough to yield the theorems that are the content of the subject. Set theory offers the mathematician great economy of primitives and premises, in return for accepting rather a lot of entities unknown to *Homo javanensis.* It offers an improvement in what Quine calls ideology, paid for in the coin of ontology. It's an offer you can't refuse.

---

[15] The above example is admittedly trivial. But it easy to produce examples in which the values of far more complicated and sophisticated definite integrals are estimated by making empirical estimations of areas under curves, yielding the type of "confirmation" of the fundamental theorem discussed above.

> The price is right; the benefits in theoretical unity and economy are well worth the entities.[16]

One may wonder why Lewis would give the above "cost-benefit" justification when he could have used the neater Argument T in support of his Realism (especially, given that he accepted the premises of Argument T). As I shall explain below, Lewis also gave a "cost-benefit" argument to justify his belief in possible worlds, so it is possible that he wanted to use his "cost-benefit" argument for belief in sets, because it could then serve as a sort of model for the kind of justification he was to give for his belief in possible worlds. Perhaps he thought that the "cost-benefit" justification for possible worlds would gain in plausibility from the logical form it shared with the justification he gave for the case of sets — its attractiveness would be enhanced by the aura of respectability surrounding the mathematical case since: (1) there was already a strong inclination among many philosophers to believe in sets; and (2) a similar sort of justification for belief in sets had been given by the highly respected philosophers Quine and Putnam.

In any case, here is how Lewis argued for the existence of possible worlds. He proposed that, by assuming the existence of possible worlds, the philosopher can reap significant theoretical benefits: by including possible worlds in our ontology, we "find the wherewithal to reduce the diversity of notions we must accept as primitive, and thereby to improve the unity and economy of theory that is our professional concern — total theory, the whole of what we take to be true.".[17] Lewis believed that "the price was right" — that the theoretical benefits of assuming the existence of possible worlds would outweigh the cost of taking on such an ontology, so that belief in possible world would be justified.[18] Thus, we can see, again, how Lewis's views about possible worlds was influenced, in its development, by his views about mathematical objects.

In what follows, I shall consider how certain crucial beliefs about the metaphysics and epistemology of mathematics that underlie Lewis's Modal Realism also play key roles in Alvin Plantinga's views about modality.

2 PLANTINGA'S THEORY OF POSSIBLE WORLDS

The Modal Realist Alvin Plantinga accepts a logical analysis of the modal operators in terms of possible worlds, as did Lewis, according to which:

---

[16] [Lewis 1986, 3-4]. For a more detailed discussion of this argument, see [Chihara, 1998], Chapter 3, Section 3. I should mention that the above argument does not constitute Lewis's sole justification for his belief in mathematical objects. In [Lewis, 1979, 182], Lewis gave a strange "ordinary language" argument to justify his belief in possible worlds. This earlier argument has received a great deal of criticism (for references, see [Chihara, 1998, 86, fn. 22].

[17] [Lewis, 1986, 4].

[18] [Lewis, 1986, 4]. For a critical appraisal of Lewis's reasoning on this point, see [Chihara, 1998, Chapter 3, Section 8].

it is possible that ϕ
is true iff
there is a possible world in which ϕ is true.

However, Plantinga's understanding of the consequent of the above biconditional differs considerably from Lewis's: for Plantinga, possible worlds are not the sort of things that Lewis believed them to be. Plantinga's possible worlds are abstract entities: more specifically, they are like incredibly complicated propositions that are so structurally complex that what they express could not possibly be comprehended by any finite mind. In what follows, I shall briefly sketch Plantinga's account of possible worlds.

The ontology of Plantinga's modal theory

(a) States of affairs

The basic entities of Plantinga's metaphysical theory of modality are *states of affairs.* States of affairs are held to be the abstract entities that such phrases as 'G. W. Bush serving on the US Supreme Court in 2004' and 'Laura Bush not being pregnant in 2004' supposedly refer to. Some states of affairs obtain, others do not. For example, the second of the two phrases I mentioned above supposedly refers to a state of affairs that obtains (since Laura Bush was never pregnant in 2004); the first one supposedly refers to a state of affairs that does not obtain. Plantinga seems to accept a sort of *abstraction axiom*: he holds that, given practically any grammatically proper gerund phrase, there is a state of affairs that the phrase denotes. (It is not clear to this writer what gerund phrases, if any, are thought by Plantinga to fail to denote existing states of affairs).

(b) Possible worlds

The concept of state of affairs is used to define what a *possible world* is. For Plantinga, a possible world is just a state of affairs that is "fully determinate" (or complete) in a way to be explained shortly. We first define two relations among the states of affairs: for all states of affairs S and S^*, S *includes* S^* iff S could not obtain if S^* did not obtain; and S *precludes* S^* iff S could not obtain if S^* did obtain. Now a state of affairs P is a *possible world* iff, for every state of affairs S, P either includes S or precludes S. To say that a possible world w is actual (or that w is an actual world) is to say that w obtains.

(c) Properties

In addition to states of affairs, Plantinga's metaphysical theory appeals to properties. For example, Socrates supposedly possessed the property of *being married to Xantippa.* What does it mean to say that Socrates possesses a property F in possible world w? It means that had w been actual, then Socrates would have possessed F.

Some properties that Socrates possess are said to be *essential properties* of Socrates. A property F is an essential property of Socrates iff Socrates possess F in every world in which he exists. A property F is an *essence* of Socrates iff F is both an essential property of Socrates and also is not an essential property of anything else. More generally, a property F is an *essence* iff there is a possible world in which there is an individual x such that: (a) F is an essential property of x; and (b) there is no world $x*$ in which some other individual exists which is such that F is an essential property of $x*$.

Plantinga's possible worlds analyses of modal notions

Two basic relationships of Plantinga's metaphysics are the relationships of *exemplification* and *coexemplification.* A property F is exemplified in a world w iff some individual would have had F had w been actual. Properties F and G are coexemplified in world w iff some individual would have had both properties had w been actual. The relationship of coexemplification is used to give some of the basic semantical definitions of Plantinga's modal system, such as the following:

$$\Diamond(\exists \mathrm{x})Fx$$

is true iff there is a possible world w which is such that for some essence E, E and F are coexemplified in w.

$$(\exists x)\Diamond Fx$$

is true iff there is some essence E which is exemplified in the actual world and, for some possible world w, E and F are coexemplified in w.

From the above, one can form some idea of how Plantinga can provide the semantics of a version of S5 quantificational modal logic.[19]

Plantinga uses his theory of states of affairs to analyze a number of well-known philosophical problems and positions such as the problem of transworld identity and the doctrine that there is no coherent distinction between *de re* and *de dicto* modal propositions.[20] It can be seen that his views about the existence of, and relationships between, the basic entities of his metaphysical system play an important role in his overall philosophical view of the world.

Some doubts about the existence of states of affairs and worlds

First of all, consider the basic entities of Plantinga's modal ontology: states of affairs. They are supposed to be abstract entities that neither exist in space nor are detectable by our most advanced scientific instruments. They are not the theoretical entities of any widely accepted scientific theory. So why should we believe in the existence of such strange entities? That there exist any such things is not a self-evident truth. And Plantinga has not put forward any empirical evidence that any such entity in fact exists. Skepticism regarding the existence of states of affairs is certainly understandable.

---

[19] See [Jager, 1982; Konyndyk, 1986] for details on how this can be done. One can find a critical discussion of these two works in [Chihara, 1998, 31–6].

[20] One can find in [Davidson, 2003] many examples in which Plantinga makes use of his metaphysical system in his theorizing about central issues in the philosophy of modality. Judging from this collection of essays, it would seem that Plantinga was a philosopher who, early on, hit upon some philosophical tools for carrying out fruitful conceptual analyses of modal reasoning of all sorts. Unfortunately, Plantinga was more concerned about exploiting the tools he had discovered than with putting his ontological theory on a firm foundation.

There is a more general reason for questioning Plantinga's postulation of states of affairs. Throughout the history of philosophy, philosophers have spun out metaphysical theories populated by strange metaphysical entities, the postulation of which is supported by the flimsiest of empirical or scientific grounds. Today, all such theories are viewed with skepticism.

Besides, even if one accepts the existence of states of affairs, why should one believe in the existence of possible worlds? Plantinga was convinced by Pollock of "the nontriviality of the claim that there is at least one possible world".[21] So a proof of the claim in question would be significant.

Why might one doubt the existence of any possible worlds, even if one were willing to accept both the existence of states of affairs and the thesis that every state of affairs has a complement? The amount of structure that a possible worlds must contain is absolutely mind-boggling. Take the actual world. Absolutely everything that exists is supposedly in this world. Not only that, every fact about every such thing is also in the world. Furthermore, Plantinga implies that every possible fact about every possible thing is also in the actual world. He tells us that every possible world is in every other possible world[22] Thus, every possible world contains the structure of every other possible world, which contains the structure of every other possible world, and so on ad infinitum. Why should we suppose that any such incredibly complex entities exist?

Plantinga's attempted proof of the existence of worlds

Plantinga tried to give a proof for the existence of at least one possible world, utilizing the following premises:

1. There are states of affairs.
2. Every state of affairs has a complement.
3. Housdorff's maximal principle.[23]
4. [Giant set existence] Given any possible state of affairs S, there exists a set whose members are all those possible states of affairs that include S.

---

[21] See [Plantinga, 1985, 327–9].

[22] See [Chihara, 1998, 123–4], for additional discussion of the structure of Plantinga's possible worlds.

[23] Housdorff's maximal principle, as formulated in standard versions of set theory, is equivalent to the Axiom of Choice.

5. [Conj] Given any set S of states of affairs, there exists a state of affairs that obtains iff every member of S obtains.

6. [Quasi-compactiness] For any set of possible states of affairs α, if α has a maximal linearly ordered subset, then α has a maximal linearly ordered subset that has the quasi-compactness property.[24]

Critical comments about the proof

For reasons that will become apparent shortly, I feel that it is not necessary for us to examine the details of Plantinga's proof.[25] Instead, let us consider some of its premises.

Start with Quasi-compactness. Plantinga does not justify using this principle, nor does he even suggests any plausible reasons for accepting it. Is it an obvious truth? Evidently, Plantinga once thought so. But Christopher Menzel has given, in detail, grounds for rejecting the principle.[26] In any case, Plantinga later became convinced by Menzel that he too should abandon the principle and, with it, his proposed proof. We can conclude that this attempt to develop a convincing proof for the existence of possible worlds is a failure.

Plantinga's use of Housdorff's principle in this proof is especially significant. Given that the principle is a standard principle of set theory — one that is generally accepted by set theorists all over the world — his use of the principle may appear perfectly reasonable. However, including this principle as a premise in his proof, along with the clearly metaphysical assertions [1], [2], [4], and [5], shows that he also regards Housdorff's principle as a metaphysical truth, clearly indicating Plantinga's acceptance of the central assumption.

It is easy to see why I classify both Plantinga and Lewis as Mathematical Realists. They both regard the existential theorems of set theory as metaphysical truths that tell us what things exists in the way many scientific laws and principles expressed in standard text books tell us what things exist in the physical universe.

Another proof based upon principles Plantinga accepted

Although Plantinga has failed to provided his readers with anything remotely like a satisfactory proof of the existence of possible worlds, it is possible to construct a much simpler proof of possible worlds from premises that he clearly accepted — a proof that does not rely upon either Quasi-compactness or Housdorff's principle. Instead, this proof proceeds from the *Ausonderungs* axiom of Zermelo-Fraenkel set theory and the following metaphysical principles all of which Plantinga clearly accepted at the time he gave his own proof:

---

[24] What Plantinga means by a 'maximal linearly ordered subset' is a set that is linearly ordered by the proper inclusion relation and that is maximal in the sense that there is no linearly ordered subset that properly includes it. A set S has the "quasi-compactness property" iff it has the following property: S is possible iff every finite subset of S is possible. A set of states of affairs is possible iff it is possible that all its members obtain.

[25] The reader can find a sketch of the proof in [Chihara, 1998, 122–4].

[26] [Menzel, 1988].

1. States of affairs exist.
2. Every state of affairs S has a complement S'.
3. If S is a state of affairs, then necessarily either S or S' obtains.
4. Given any possible state of affairs S, there exists a set whose members are all those possible states of affairs that include S.
5. Given any set S of states of affairs, the conjunction of that set exists.[27]

Problems with the proof

Now it is a relatively simple matter to prove that the above group of "safer" principles used in this second proof is inconsistent with the axioms of set theory. Thus, it can be seen that the system of principles Plantinga used in his theorizing about modality runs into conflict with classical set theory. At one time, Plantinga evidently thought he could obviate this sort of problem by rejecting Cantor Theorem.[28] However, it is hard to see how such a resolution of the problem could be reconciled with Plantinga's overall Platonic view of set theory,[29] and evidently Plantinga never developed this idea.

Can it be repaired?

Later, Plantinga became convinced by Menzel that he should abandon premise [4].[30] Unfortunately for Plantinga, dropping this premise from the list of principles he accepts will not save his system from conflict with the standard axioms of set theory. It can be shown that, if one accepts the other principles of Plantinga's metaphysical system, one can still derive a contradiction with the theorems of set theory.[31] Furthermore, it is hard to see how one could trim the system so as to yield a theory that, on the one hand, is consistent with set theory and, on the other hand, will provide Plantinga with the kind of modal framework he needs to achieve the analysis of modality that he has for so long championed. The problem is that he has accepted a number of ontologically extravagant principles concerning sets, properties, states of affairs, and propositions, all of which are logically interconnected in a way that results in a system that is not only inconsistent with set theory, but also so complicated that it is hard to see how one could develop a reasonable model theory for it.[32] I conclude that the ontological system Plantinga works with is a mess. I thus regard as remote, not only the probability of his coming up with a reasonably secure and convincing proof of the existence of possible worlds, but even the prospects of his giving his present metaphysical system the kind of relative model theoretic security that standard systems of set theory have today. *I thus view with considerable skepticism practically all of the substantial conclusions about modality he has arrived at on the basis of his metaphysical system.*

---

[27] See [Chihara, 1998, 125–6], for the details of this proof.

[28] See [Chihara, 1998, 128], for the relevant references.

[29] The rejection of Cantor's Theorem is generally linked to severely limiting the Abstraction Axiom along constructivistic lines such as proposed by Predicativists.

[30] This, I learned from a letter he wrote to me dated 11 September 1991.

[31] See [Chihara, 1998, 125–6].

The Mathematical Realism implicit in Plantinga's Modal Realism

It is now time to take up a more fundamental feature of Plantinga's Modal Realism. Recall that underlying Plantinga's arguments that there are possible worlds is his acceptance of the central assumption and his Mathematical Realism. This aspect of Plantinga's ontological theory may not have appeared at all controversial to Plantinga. After all, a great many philosophers seem to accept the central assumption and, of all the many different sorts of entities that Plantinga accepts, mathematical objects are probably accepted by the greatest number of philosophers. And it is not just philosophers who believe in the existence of mathematical objects: so also do many mathematicians and logicians. Furthermore, famous mathematicians, logicians, and philosophers have supported their belief that there are mathematical objects by setting forth a variety of philosophical arguments — a few of which have won, if not widespread approval, at least very serious consideration.[33] It should be noted that the Mathematical Realism implicit in Plantinga's account of modality is not just a feature of his failed proof of the existence of possible worlds; it figures in many of his writings on modality and is clearly a deeply entrenched part of his overall philosophical outlook, as the reader can easily verify.[34]

Another aspect of Plantinga's Mathematical Realism is revealed in his book *Warranted Christian Beliefs*, where he tells us that, among our cognitive faculties, one of the most important is *reason*:

> Taken *narrowly*, reason is the faculty or power whereby we form *a priori* beliefs, beliefs that are *prior* to experience or, better, independent, in some way, of experience. These beliefs include... simple truths of arithmetic and logic, such as *2 + 1 = 3* and *if all men are mortal and Socrates is a man, then Socrates is mortal.* They also include... more

---

[32] For additional details concerning my doubts about the coherence of Plantinga's metaphysical system, see [Chihara, 1998, 129–411].

[33] The most widely discussed of these arguments is Quine's "indispensability argument". I discuss this argument in great detail in [Chihara, 2004, Chapter 5].

[34] See, for example, [Plantinga, 2003, 48], where he explains what *books* and *possible worlds* are using set theory. It is clear that a realistic attitude toward sets thoroughly suffuses the web of ideas expressed in these passages. There is not even a shadow of a doubt expressed about the existence of the sets being presupposed in the reasoning.

> controversial items such as the belief that there are... abstract objects.[35]

Notice that Plantinga classifies the proposition that $2 + 1 = 3$ as a truth. From this quotation, one can see that Plantinga believed, as did Kant before him, that we have *a priori* knowledge of mathematical truths. Furthermore, we even find a suggestion that we can come to know *a priori* that there are abstract objects such as numbers.

Like Lewis's view of modality, Plantinga's metaphysical and epistemological beliefs about mathematics form an essential part of the support for his Modal Realism. Indeed, it is hard to imagine Plantinga putting forward the kind of proof for the existence of possible worlds described above had he not understood proofs in set theory in the Platonic way he clearly does. One can see how he has modeled his reasoning about possible worlds on the pattern of set- theoretical proofs. In a way that brings to mind the Kantian strategy for responding to Hume, the widely held belief that set theorists are able to provide us with knowledge of the abstract entities of the mathematical universe undoubtedly emboldened Plantinga to imagine that he too could obtain remarkable ontological results about possible worlds by reasoning as set theorists do.

The widespread acceptance of both the central assumption and also the realistic view of mathematics is undoubted one reason both Lewis and Plantinga are so insouciant about postulating and referring to such things as properties, states of affairs, and propositions — abstract entities that are, in logical features, so much like those of sets that their postulation may appear, to anyone who already believes in sets, to be a natural and unobjectionable step. These metaphysical beliefs about mathematics help clear the way for the lush ontology they proposed for their modal theories. In both cases, their Mathematical Realism serves to enhance the acceptability of the other kinds of metaphysical entities that their respective modal theories require. Thus, casting doubt on their philosophical beliefs about mathematics would also serve to cast doubt upon their versions of Modal Realism.

3 AN EXAMINATION OF ARGUMENT T

In this section, I shall attack the Realistic view of mathematics upon which the above two Modal Realists base so much, by showing the dubiousness of Argument T.[36] This will be done, in particular, by undermining the grounds given earlier for accepting its principal premise (the central assumption). I shall begin by showing how a mathematical theorem can be used (or "applied") in ordinary everyday situations and in science to infer a factual empirically testable conclusion from some factual empirically testable premises, even when the theorem itself is not assumed to be a true statement. It will be shown that it is not the *truth* of the theorem that is crucial in such cases of applications of mathematics but rather another feature of the theorem which I call its "structural content".

---

[35] [Plantinga, 2000, 146–7]. What he writes in another place in the book sheds further light on his views about *a priori* knowledge: " Among the things we know, some are *self-evident*... [where, roughly] a proposition is self-evident if it is so utterly obvious that we can't even understand it without seeing that it is true. Examples would be propositions like *7 + 5 = 12*... And the idea is that reason, taken in this narrow sense, is the faculty or power whereby we see the truth of self-evident propositions. Of course it is also reason whereby we see that one proposition implies or entails another... (p. 113)."

[36] One might reply that Lewis's "cost-benefit" justification for belief in sets would not be undermined by such a result. However, I have already given a series of detailed objections to that justification in [Chihara, 1998, Chapter 3, Section 8].

For purposes of clarity and simplicity, I shall start with the case of a theorem of a first-order mathematical theory (although the basic ideas apply to mathematical theories expressed in ordinary languages). Thus, I shall first assume that the mathematical theory to be considered is expressed as an axiomatized deductive theory formalized in a first-order logical language (as is Zermelo–Fraenkel set theory). I shall also assume that the theory is given an appropriate sort of interpretation, so that, under this interpretation, the sentences of the theory can be regarded as making genuine assertions. The kind of interpretation I have in mind are the natural language interpretations that philosophically minded logicians have devised for this purpose: I call them "natural language interpretations" (or "NL Interpretations" for short). These are the "interpretations" of first-order languages that *philosophically trained* logicians are apt to consider when "translations" of the logical language into some natural language are seriously contemplated. These "interpretations" do more than what mathematical structures do: they not only assign the relevant sort of sets and objects to the parameters of the logical language in question, they also supply meanings or senses to the parameters. Interpretations of this sort assign to each individual constants the sense of some English name or definite description; and they provide each predicate of the language with the sense or meaning of an *English predicate*, where English predicates are obtained from English declarative sentences by replacing occurrences of names or definite descriptions with occurrences of circled numerals. When a logical language is given an NL interpretation, the sentences of the language can be regarded as expressing statements that are true or false. For example, if the interpretation **I** assigns to 'R^2' the sense of the English predicate '① is heavier than ②', and to 'a' and 'b' the senses of 'The Governor of California' and 'Gray Davis' respectively, then the sentence '$R^2 ab$' expresses the statement 'The Governor of California is heavier than Gray Davis'. (It does not express the statement 'Arnold Schwarzenagger is heavier than Gray Davis' even though what it expresses is true iff the second statement is true). I shall say that a logical language equipped with the meanings or senses provided by an NL interpretation is an *NL interpreted language.*

Suppose now that a scientist (or engineer for that matter) applies a mathematical theorem in drawing an empirically testable conclusion from a number of empirically established premises. If the scientist's reasoning is to be intelligible and rational, must the theorem used in the inference be assumed to be a true statement? Could the scientist's inferences in question be rationally justified *even when that mathematical theorem is not assumed to be true*?

As a preliminary to answering the questions of the previous paragraph, I suggest that, independently of whatever "ordinary" meaning a theorem of a mathematical theory may have, it can be understood to have a kind of content, its "structural content, which tells us what must hold in structures of a certain sort — the particular sort in question depending upon the kind of structures with which the theorem prover was concerned. Thus, to consider a specific case, suppose that ϕ is a theorem of an axiomatized first-order mathematical theory M. Then we know from logic that ϕ must be satisfied by every first-order structure that satisfies the axioms of M. We can then infer the truth of the *structural content* of ϕ, which can be expressed

> Every model of M must also be a model of ϕ.[37]

Obviously, the structural content of ϕ can be seen to hold without having to establish the truth of ϕ under its "ordinary interpretation".

It will now be argued that, when typical applications of M are considered, *there is no need to assume that the theorems of M are true statements; it is sufficient that the structural content of the theorem hold.* To see how this could be, let us now consider the case in which M is a specific axiomatized first-order mathematical theory, viz. Peano arithmetic (henceforth '**PA**').[38]

Let the language of **PA** be extended to a many-sorted language so that there are quantifiers for natural numbers and other quantifiers for empirical objects, such that, *when the new language is given an appropriate sort of NL interpretation, empirical statements containing arithmetical terms are expressible in the language.*[39] Consider now what entities must be assigned by an NL interpretation to the natural number terms in order for some sentences of the many-sorted language being discussed to express true empirical statements. Any NL interpretation of the target language must make assignments to both the arithmetical and non-arithmetical constants of the language, as well as to the predicates and terms that take both arithmetical and non-arithmetical variables. Thus, an appropriate domain for the arithmetical quantifiers must be determined and referents must be selected (with respect to this domain) for all the purely arithmetical constants, terms, and predicates; a domain for the non-arithmetical quantifiers must also be determined and referents must be selected (with respect to this latter domain) for all the purely non-arithmetical constants, terms, and predicates; while at the same time, referents must be selected (with respect to both of the above domains) for the terms and predicates that take both arithmetical and non-arithmetical variables.

---

[37] For the case in which M is not a first-order theory, one could reason (roughly) as follows: Since ϕ is a theorem of M, ϕ must be provable from the axioms of M, and hence it must be a consequence of these axioms. We can infer that any model of these axioms must be a model of ϕ.

[38] I do not wish to leave the reader with the impression that my reasoning must be restricted to only cases in which the mathematical theorem applied is a theorem of an axiomatized first-order theory. I have chosen to discuss the reasoning within such a framework for the sake of simplicity and clarity.

[39] See [Enderton, 1972, Section 4.3] for a rigorous discussion of many-sorted logic.

Suppose that there is some sentence of this language that can be reasonably translated:

(1) There are 3 horses in the barn

What must be assigned to the constant term that is being translated here by the numeral '3' in order that the sentence be true? It has been convincingly maintained by a number of philosophers classified as "Structuralist", such as Paul Benacerraf, Michael Resnik, Stewart Shapiro, and Geoffrey Hellman, that it does not matter what specific entity '3' denotes, so long as the entities that are denoted by the natural number terms (the entities that form the domain of the arithmetical quantifiers):

> (a) form an ω-sequence (or what Bertrand Russell called a "progression"[40]);

and

> (b) the elements of the progression are so related to one another that the member of the progression denoted by '3' is appropriately related to the other members of the domain (it must be, so to speak, "the successor of the successor of the successor of" the member denoted by '0').

Thus, whether one takes the natural numbers to be the Zermelo ordinals

$$\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots$$

or the von Neumann ordinals

$$\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots$$

the truth value of the empirical statement in question will not be affected, since it will be true iff there is a one-one correspondence between the totality of *horses in the barn* and the initial segment of either one of the above sequences consisting of those members of the sequence that precede the entity denoted by '3'.[41] It is easy to see that the truth conditions of (1), and hence the empirical content of (1), is not dependent upon what specific entity is assigned as the denotation of '3'. In so far as what is of concern is the empirical content of such sentences, it does not matter what entities we pick to be the natural numbers — any progression will do: it is only how these entities are interrelated that counts. In other words, regarding the referents needed for the arithmetical terms, *so long as the NL interpretation selects a progression as the natural numbers and fixes appropriate members of the progression as referents for the arithmetic constants, the truth*

---

[40] [Russell, 1920, 8]. There are, of course, nonstandard models of Peano arithmetic, but for reasons of avoiding unnecessary complications, I shall restrict the above thesis to considerations of only "standard" models.

[41] This was implicitly argued by Benacerraf in (Benacerraf, 1965).

values of the empirical statements will be correct. Any NL interpretation that satisfies the above conditions regarding both the empirical and the arithmetical parts of the language, I shall call an "empirically adequate interpretation", and I shall refer to the above thesis about the referents of the arithmetical terms and the arithmetical quantifiers, the "Structuralist Thesis".

Now let **I** be an NL interpretation of this language such that each member of G would naturally be translated into true English empirical sentences: we can imagine that the non-arithmetical constants of the language have been assigned appropriate elements with respect to some domain of physical entities.

What about the arithmetical part of the language? It is usual for the domain of the arithmetical part to be specified in essentially the following way:

1. the domain is to consist of the non-negative integers;
2. a specific individual constant is to denote the number 0;
3. a specific unary successor operator is to denote the number-theoretic function that yields as value, for any argument α, the immediate successor of α;
4. a specific binary addition operator is to denote the binary operation that yields, for any ordered pair of non-negative numbers $< \alpha, \beta >$, the sum of α and β;
5. a specific binary multiplication operator is to denote the binary operation that yields, for any ordered pair of non-negative numbers $< \alpha, \beta >$, the product of α and β.[42]

Now as philosophers, we should ask if the above specification of the referents for the arithmetical part of the language really yield a model of **PA**. Have we actually specified genuine denotations for all the nonlogical arithmetical constants of the language? No, we haven't — at least if the nominalists are correct. Since nominalists do not think there are such entities as non-negative integers, if they are right, the above specification of the referents of the arithmetical constants of the target language has not succeeded in securing the required domain and referents for all the nonlogical constants of the language (anymore than one succeeds if one specifies that the domain is to consist of all the unicorns that are to be found in California). In that case, the NL interpretation specified above has not provided each arithmetical sentence of the language with a meaning resulting in a true or false sentence. Thus, the axioms of **PA,** expressed in the language so interpreted, could not be correctly said to be true sentences. If the nominalists are right, then, the above NL interpretation does not succeed in making the theorems of **PA** true statements.

---

[42] The specification I have given above is routinely given. For example, see [Mendelson, 1987, 121], where he goes on to say: "This interpretation is called the *standard interpretation* or *standard model*". A similar specification of a model of PA is given by Enderton in his [Enderton, 1972, 174].

What about the sentences of the language that are translated into straightforward empirical statements such as 'There are 3 apples on the table'? If the NL interpretation does not secure a referent for the arithmetical constant that is translated by '3', how can such a sentence be regarded as true? Well, we look at the surface of the table and count all the apples on it: one, two, three. In this case, we would all regard the sentence as true — without concerning ourselves about what *particular entity* has been assigned as the referent of the word 'three'. It doesn't matter what thing (if any) the word is supposed to denote, since, by the Structuralist Thesis, any initial part of a progression would fill the role needed for the natural numbers — even the sequence of English words 'one', 'two', 'three', 'four', . . . would do. Obviously, the empirical statements only involve the natural numbers tangentially (if at all), and in ordinary situations, we do not concern ourselves with the question of what particular referents have been or should be selected for the natural numbers. Thus, whether or not the NL interpretation selected has succeeded in picking out actual (existing) entities as referents of the natural number terms is usually not even considered in determining the truth values of such empirical statements. Even the most committed nominalists would not object to our taking the empirical statements of the above language as having truth values.

The situation is very different, however, in the case of the sentences of **PA**: those sentences are supposed to be directly and crucially *about the natural numbers.* If no referents have been secured for the referential terms of those sentences, then it is hard to see how the sentences can be correctly said to be *about anything* and hence to be either true or false.

Even if the above reasoning is sound, it may still be wondered how, if none of the theorems of **PA** have truth values, it can be correct to infer some empirical sentence θ from the set of empirical sentences G on the basis of some theorem ϕ of **PA**. If one does not take the theorem ϕ to be true, how can one rationally make use of ϕ to infer θ from G? Recalling the arguments for the central assumption discussed earlier, can one have any faith in empirical conclusions drawn from ϕ, if ϕ is not known to be true?

With the aim of answering these questions, let us suppose that **I** is an NL interpretation of this language that satisfies the set of sentences G in the way described above. Let us also suppose that empirical sentence θ is deduced from G using theorem ϕ as a premise. In the following, I shall make no assumption about whether or not the NL interpretation has in fact secured referents for the domain of **PA** and hence whether or not the sentence ϕ has a truth value, since from the point of view of the nominalist, the axioms and theorems of **PA** (and hence ϕ itself) are not true statements. Still, we know that, since ϕ is a theorem of **PA**, its structural content must hold, that is, ϕ must be true in all models of **PA**. So we can infer that ϕ must be true in all empirically adequate models of **PA** that are "consistent with" **I**.

The above consistency condition needs a bit of explanation, since it is brought in only to cover unnatural and probably perverse NL interpretations of the language.

Suppose that the NL interpretation **I** had assigned the Eiffel Tower to some non-arithmetical constant α and that what G said about α (under this interpretation) blocked α from being a natural number. Then, any empirically adequate model of **PA** that had the Eiffel Tower in its domain would conflict with interpretation **I**. Such an empirically adequate model would, in a sense, be inconsistent with the interpretation.

Still, if ϕ is true in *all* models of **PA**, then any empirically adequate model of **PA** that is not "inconsistent with" **I** in the above way must satisfy ϕ. Thus, we can infer from the above discussion that ϕ must be true in all empirically adequate models of **PA** that are "consistent with" **I**.

Now it is possible to construct an NL interpretation **K** of the language that is identical to **I** except for what it assigns to the domain of the arithmetical quantifiers of the language and to the arithmetical constants of the language, such that **K** will be an empirically adequate model of **PA** consistent with **I**.[43] Since **K** is a model of **PA** and since the structural content of ϕ is true, **K** must satisfy ϕ. From the Structuralist Thesis, **K** must also satisfy G. Since θ is a consequence of G and ϕ, **K** must satisfy θ. And since **K** differs from **I** only in what these two interpretations assign to the arithmetical domain and to the arithmetical constants, it follows by the Structuralist Thesis that θ must be true under **I**. That is, θ must be a true sentence. It can thus be concluded that the deduction in question can be rationally justified without having to assume that ϕ is a true statement.

The above case of the use of a theorem of **PA** may appear very artificial since the reasoning is carried out in a rigid, highly restricted formal language that concerns only a relatively simple mathematical structure. So **I** would now like to consider briefly the case of the use of a mathematical theory expressed in a natural language and involving a more complex kind of structure (analysis) than the intended models of **PA**.

A theory of light rays

Imagine that a scientist, on the basis of her observations of light and shadows, coupled with a number of experiments involving light sources, mirrors, and opaque screens that have pinholes, constructs a mathematical theory[44] in which 'source', 'observer', and 'point of reflection' are used to refer to points in Euclidean 3-space with the standard metric. In the theory, a continuous curve connecting any two points a and b is called a 'path from a to b'. The term 'path of a light ray from point a to point b' is used to refer to a special kind of path from a to b. By definition, a "path of a light ray from a to b" is the shortest path from a to b.

A *light reflection instance* consists of a path **f**, a plane **p**, and points a, b, and c such that:

---

[43] Such an NL interpretation is constructible using the author's Constructibility Theory, as is shown in [Chihara, 2004, Chapter 7].

[44] The theory is presented in detail in "The Existence of Mathematical Objects" (forthcoming in a collection of essays to be published by the Mathematical Association of America, edited by Bonnie Gold and Roger Simons).

> c lies on **p**; a and b are not on **p**, but are both in the same half space bounded by **p**; and **f** is the union of a path of a light ray from a to c and a path of a light ray from c to b, all the parts of **f** being in the same half space bounded by **p**.

Such an **f** is called a 'path of reflection (relative to **p**) from a to b'; c is called a 'point of reflection' of **f**. The term 'path of reflection' functions as a parameter (as do the nonlogical constants of a first-order theory): in all models of the theory, the term refers to paths satisfying the following axiom:

> A path of reflection (relative to **p**) from a to b is the shortest possible path from a to b that contains a point of reflection in **p**.

Not surprisingly, when the above parameter is suitably interpreted, the theory can be understood to describe features of a simplified model of idealized light rays traveling in a homogeneous medium. Thus interpreted, a theorem can be proved which can be understood to say:

> Any path of reflection will be such that the path's angle of incidence equals its angle of reflection.

We can see how this theory can be applied to actual physical situations involving the behavior of light, by regarding paths of actual light rays to be approximately that of paths of light rays from a source to an observer (as characterized above). The physical situation would have to be ones in which the light analyzed would be traveling in "homogeneous media" — say, reasonably clear (fogless, smokeless) air — and being reflected by a flat object, such as a mirror, approximating a plane in certain geometrically relevant ways. In such a situation, we would have, in an idealized way, a kind of *empirical model* of the theory. The paths of light rays in such homogeneous media are in fact very close to straight lines (for most practical purposes) so that the path of a light ray from point a to point b can be taken to be the shortest distance between a and b. Furthermore, the geometric properties of flat mirrors can be made to be very close (at least for many practical purposes relating to how light is reflected by mirrors) to those of planes, so that paths of light rays reflecting off such a mirror would be very close to what is described in the principal axiom.

Now suppose that the parameter 'path of reflection' is left uninterpreted. In that case, any theorem containing the parameter does not assert anything (anymore than does the sentence of first-order logic 'Fa' when the constants in it are not interpreted), and hence it would not assert anything true or false about the actual world. However, regardless of what incomplete meaning such a theorem will retain, it will still have a structural content which can be shown to hold. Hence, any model of the above theory would have to be a model of the theorem. Then, since the empirical model described above satisfies the axioms (at least as an approximation), the "model" would have to satisfy the above theorem. This will allow us to conclude that a light ray striking a mirror at a 35 degree angle will reflect off the mirror at (approximately) a 35 degree angle.[45]

We can now see how a mathematical theorem could be legitimately applied even when the theorem used is not assumed to be a true assertions.[46] Thus we can see how one can undermine the Resnik premise (discussed earlier) that we are justified in drawing conclusions from and within science only if we are justified in taking the mathematics used in science to be true.

The fundamental theorem reexamined

The above example provides the means of understanding how there can be the kind of "empirical verification" of the Fundamental Theorem discussed earlier, without having to assume that the theorem is a true statement. We have: on the one hand, a purely mathematical theory (analysis) in which the Fundamental Theorem is proved; and on the other hand, an empirical theory (say, about the printed squares of engineering paper) according to which certain kinds of entities are definitely related in ways that approximate, more or less accurately, the ways that various elements of structures are asserted to be interrelated by the structural contents of the axioms of analysis.[47] It is natural to speak, in this context (as in the previous example), of a sort of "empirical model" — one that is structurally identical, given certain idealizations, to the models (or sub-models) of the mathematical theory.[48] Then, the fact that the theorems of the mathematical theory are true of all the models of the theory explains why the Fundamental Theorem would hold of the "empirical model" used to carry out the "empirical verification". Thus, it is not the actual truth of the Fundamental Theorem that is "empirically verified" but rather only its structural content. It can be seen that this kind of "verification" no more presupposes the actual truth of the Fundamental Theorem itself than does an "empirical verification" of the previously discussed law of reflections of light rays presuppose the truth of the theorem of the previously discussed mathematical theory of light. In both cases, all that is really presupposed is the truth of the structural contents of the respective mathematical theorems in questions.[49]

Some lessons to be learned

The above discussion leads naturally to the view that the theorems of practically any mathematical theory need not be true to be applied in science and in everyday situations. It undermines the strong conviction that is so prevalent among philosophers that the theorems of mathematics are factual statements that have to be true of the world (the central assumption) in order that they be applied in science and engineering.[50]

---

[45] Since we can infer that the angle of reflection of the path of the actual light ray is (approximately) equal to its angle of incidence.

[46] For an indication of how the above theory can be developed into a more complex and versatile theory, see [Maki and Thompson, 1973, Sec. 2.3.4].

[47] Cf. the example discussed in [Chihara, 2004, 285–6].

[48] Cf. [Chihara, 2004, Ch. 9, Sec. 12, especially fn. 59].

[49] For a detailed explanation of the underlying point being made here, see [Chihara, 2004, 9, Sec. 12].

There is a tendency among philosophers and mathematicians alike to think that to deny that mathematical theorems are true statements is to imply that both theorems and non-theorems alike fail to provide us with any useful information — an implication that seems to be contradicted by the obvious fact that scientists and engineers find many mathematical theorems to be extremely informative. What the above discussion brings out is that, even if a theorem is not true, it can still have a kind of content (its structural content) that conveys valuable information to scientists and engineers.[51]

4 CONCLUDING COMMENT

I am inclined to call the central assumption a "dogma" because it is held by so many philosophers without any real investigation into the question at issue. Just how deeply Lewis and Plantinga had seriously investigated the question of the truth of mathematics before accepting the central assumption is not known by me. But judging from their writings, I doubt that either of these philosophers had ever seriously questioned this assumption. In any case, noting that these Modal Realists, in effect, utilized the central assumption at crucial points in their philosophizing about the metaphysics of modality (without even mentioning the assumption about mathematics they were making) points to a serious area of vulnerability of their ontological systems. The above discussion shows how one can reasonably question their fundamental assumptions about the nature of mathematics and thereby cast doubt upon the fragile and largely unexplored underpinnings of their metaphysical views about modality.

---

[50] It might be objected that my refutations presuppose the truth of mathematics. For example, does not the above reasoning concerning **PA** makes use of some elementary theorems of first-order model theory — a theory that relies upon set theory? My response to this objection is to appeal to the Constructibility Theory of my book [Chihara, 1990]. The Constructibility Theory was developed to be a theory that neither makes reference to, nor presupposes, the existence of mathematical entities but that can still serve as a logical framework for the development of much mathematical theory, including model theory. The model theory needed above can easily be developed within the Constructibility Theory.

[51] Here, it might be objected that such apparent reference to structures runs counter to my skepticism about mathematical objects. For if the nominalist is correct, how can there be any models of mathematical theories? Aren't structures abstract entities on a par with sets and properties? Furthermore, if there aren't such things as stuctures, how can the structural content of mathematical theorems have any real significance? Here again, the Constructibility Theory comes to the rescue by providing the nominalist with "constructability structures" — something that the nominalist can use to make sense of applications of mathematical theorems in science and to answer the above questions. For more details on this point, see [Chihara, 2004, Chapter 8].

ACKNOWLEDGEMENTS

Material for this paper was delivered in a series of three lectures I gave at the *Scuola Normale Superiore* in Pisa in the third week of April of 2004. I am grateful to those who attended the lectures and contributed many thoughtful comments on the earlier material.

BIBLIOGRAPHY

[Benacerraf, 1965] P. Benacerraf. What numbers could not be. *The Philosophical Review, 74*, 47-73, 1965.

[Burgess and Rosen, 1997] J. Burgess and G. Rosen. *A Subject With No Object: Strategies for Nominalistic Interpretations of Mathematics.* Oxford: Oxford University Press, 1997.

[Chihara, 1990] C. S. Chihara. *Constructibility and Mathematical Existence.* Oxford: Oxford University Press, 1990.

[Chihara, 1998] C. S. Chihara. *The Worlds of Possibility: Modal Realism and the Semantics of Modal Logic*. Oxford: Oxford University Press, 1998.

[Chihara, 2004] C. S. Chihara. *A Structural Account of Mathematics.* Oxford: Oxford University Press, 2004.

[Davidson, 2003] M. Davidson, ed. *Essays in the Metaphysics of Modality.* New York: Oxford University Press, 2003.

[Enderton, 1972] H. Enderton. *A Mathematical Introduction to Logic.* New York: Academic Press, 1972.

[Jager, 1982] T. Jager. An actualist semantics for quantified modal logic. *Notre Dame Journal of Formal Logic*, 23, 335–349, 1982.

[Konyndyk, 1986] K. Konyndyk. *Introduction to Modal Logic.* Notre Dame, Ind.: University of Notre Dame Press, 1986.

[Lewis, 1979] D. Lewis. Possible worlds. In M. J. Loux (Ed.), *The Possible and the Actual*, pp. 182-189. Ithaca: Cornell University Press, 1979.

[Lewis, 1986] D. Lewis. *On the Plurality of Worlds.* Oxford: Basil Blackwell, 1986.

[Lewis, 1991] D. Lewis. *Parts of Classes.* Oxford: Basil Blackwell, 1991.

[Maki and Thompson, 1973] D. P. Maki and M. Thompson. *Mathematical Models and Applications.* Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

[Mendelson, 1987] E. Mendelson. *Introduction to Mathematical Logic* (3rd ed.). Monterey: Wadsworth & Brooks/Cole Advanced Books, 1987.

[Menzel, 1988] C. Menzel. On an unsound proof of the existence of possible worlds. *Notre Dame Journal of Formal Logic*, 30, 598 - 603, 1988.

[Plantinga, 1985] A. Plantinga. Replies to my colleagues. In J. Tomberlin & P. van Inwagen (Eds.), *Alvin Plantinga*, pp. 313-396. Dordrecht: D. Reidel, 1985.

[Plantinga, 2000] A. Plantinga. *Warranted Christian Belief.* New York: Oxford University Press, 2000.

[Plantinga, 2003] A. Plantinga. World and essence. In M. Davidson (Ed.), *Essays in the Metaphysics of Modality*, pp. 46-71. New York: Oxford University Press, 2003.

[Resnik, 1997] M. Resnik. *Mathematics as a Science of Patterns.* Oxford: Oxford University Press, 1997.

[Resnik, 1998] M. Resnik. Holistic mathematics. In M. Schirn (Ed.), *The Philosophy of Mathematics Today*, pp. 227-246, 1998. Oxford: Oxford University Press.

[Russell, 1920] B. Russell. *Introduction to Mathematical Philosophy* (Second ed.). London: Allen & Unwin, Ltd, 1990.

FREE LOGICS

John Nolt

1 THE BENEFITS OF FREEDOM

In classical predicate logic each singular term must denote some member of the quantificational domain. Free logics annul this requirement. Their quantifiers, like classical quantifiers, are interpreted over a specified domain, but their singular terms may denote either objects outside that domain or nothing at all. If we regard the quantificational domain as the set of existing things, free logics are logics whose singular terms need not denote existents. Just as modern syllogistic widens the applicability of Aristotelian syllogistic by accommodating general terms (e.g., 'celestial sphere' or 'infinitely large rational number') that apply to no existing thing, so free logic widens the applicability of classical predicate logic by accommodating singular terms (e.g., 'the firmament' or '1/0') that denote no existing thing.

Application of classical predicate logic to statements whose singular terms are *not* known to denote existing things may beg important questions. Rendering 'God', for example, as g, we obtain by classical reasoning the remarkable theorem $\exists x\ x = g$. But the same reasoning also yields, alas, the existence of Godzilla. To avoid such embarrassments, classical reasoners are obliged to confirm the existence of things named before naming them.[1]

But this obligation is irksome, so irksome that even mathematicians routinely flout it — using, for example, the expression '$f(x)$' as a singular term even where f is partial and the existence of $f(x)$ unproved. They get away with this, usually, only by being discreetly inexplicit about the underlying logic — which is, in consequence, not rigorously classical.

Singular terms of dubious denotation crop up not only in mathematics but also in science ('the center of the universe', 'the ether', 'Vulcan'[2]), in history ('Homer', 'King Arthur') in metaphysics ('my mind', 'the One', 'nothingness') and just about everywhere else. As in mathematics, we may want to reason with such terms, even without assurance that they refer to existing things.

---

[1] I am assuming here, as is usual, that the existential quantifier is read "there exists." Later I'll critique this assumption.

[2] 'Vulcan' is the name given by Urbain Le Verrier — the nineteenth century astronomer who performed the calculations that led to the discovery of Neptune — to a planet which, he believed, orbited between Mercury and the sun, perturbing Mercury's orbit. He and others searched unsuccessfully for Vulcan, which in fact does not exist. Mercury's behavior was later explained by (and helped to confirm) the general theory of relativity.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

Free logics enable us to do so by disallowing inferences whose validity depends on the presumption that singular terms denote members of the quantificational domain. Existential generalization:

$$A(t/x) \vdash \exists x A$$

is a case in point. For if t does not denote an object in the quantificational domain then the truth of $A(t/x)$ may not guarantee that this domain contains an object that satisfies A. Without existential generalization, the assumption $g = g$ no longer implies $\exists x\ x = g$; thus we may use the name 'God' — or 'Godzilla' — without existential presumption.

Universal instantiation (specification):

$$\forall x A \vdash A(t/x)$$

is likewise invalid; for even if every object in the domain of the universal quantifier satisfies A, if t does not denote a member of that domain, $A(t/x)$ may be false. Because free logics disallow some classical inferences but validate no inferences that were classically invalid, they are weaker than the corresponding classical logics.

To distinguish terms that denote existing things from those that do not, free logics often employ the one-place existence predicate, 'E' or '$E!$'. (I favor the latter, because it is pronounced "E-shriek.") For any singular term t, $E!t$ is true if t denotes an object in the domain of the existential quantifier, false otherwise. For free logics with identity, '$E!$' is definable thus:

$$E!t =_{df} \exists x x = t^3$$

and hence need not be taken as primitive. ('$E!$' is not definable if the identity predicate is absent.[4]) Using '$E!$' we can formulate weakened versions of existential generalization:

$$A(t/x), E!t \vdash \exists x A$$

and universal instantiation:

$$\forall x A, E!t \vdash A(t/x)$$

that are valid in free logic. Here the classical presumption that singular terms denote members of the quantificational domain becomes an explicit assumption, $E!t$, that can be made or not made separately for each term t.

---

[3] There is one possible exception. If $s = t$ is truth-valueless for all referring s and nonreferring t, then (depending on the interpretation of quantifiers) $\exists x x = t$ may be truth-valueless while $E!t$ is false. See the discussion of nonvalent free logics below.

[4] This was first shown in Meyer, Bencivenga and Lambert [1982].

2 THE EMPTY DOMAIN

A consequence of the classical presumption is that the domain must contain something for singular terms to denote. It must, in other words, be nonempty. Thus classical logic treats many existential claims — not only $\exists x\ x = t$, but also, for example, $\exists x\ x = x$, $\exists x Fx \vee \exists x \sim Fx$, even $\exists x(Fx \rightarrow \forall y Fy)$ — as logical truths. But mere logic, it would seem, ought not to guarantee the existence of anything. In *Introduction to Mathematical Logic*, Russell writes:

> Among 'possible' worlds, in the Leibnizian sense, there will be worlds having one, two, three, ... individuals. There does not even seem any logical necessity why there should even be one individual.[5]

And in a footnote he adds:

> The primitive propositions in *Principia Mathematica* are such as to allow the inference that at least one individual exists. But I now view this as a defect in logical purity.

One might question Russell's regard for purity. Still, in free logics where we need no objects for singular terms to denote, why not allow the domain sometimes to be empty? Quine [1954] calls logics that admit the empty domain *inclusive*. In an inclusive logic no existentially quantified formula is valid.

Inclusive logics need not be free (though they must be if they employ individual constants), nor need free logics be inclusive. Inclusivity furthers the spirit of freedom, however, by eliminating the last vestige of existential import from general terms. In classical predicate logic, universal general terms (terms satisfied by all members of the domain) still have existential import; that is, any inference of the form $\forall x A \vdash \exists x A$ is valid. But in inclusive logics the empty domain is a counterexample to all such inferences.

Inclusive logics, whether free or not, have, however, their peculiarities. They make universally quantified contradictions — $\forall x(Fx\ \&\ \sim Fx)$, for example — contingent, since all universally quantified formulas are true in the empty domain. This includes vacuously quantified formulas, if we allow them. Hence they invalidate:

$$\forall x A \rightarrow A, \text{ where } x \text{ is not free in } A,$$

which is valid not only in classical logic but also in free logics that are not inclusive. More troublesome is the loss of some of the laws of confinement — e.g.:

$$\forall x(P\ \&\ A) \leftrightarrow (P\ \&\ \forall x A), \text{ where } x \text{ is not free in } P,$$

that are needed for the operations of prenexing (moving all quantifiers to the beginning of a formula) or purifying (moving quantifiers in as close as possible to the variables they govern). This loss invalidates many equivalences that classical logicians take for granted. The formula:

---

[5] Russell [1919], p. 203.

$$\forall x(A \leftrightarrow x = t),$$

for example, which is prominent in the theory of definite descriptions, is not equivalent in inclusive logic, as it is in classical logic, to:

$$\forall x(A \rightarrow x = t) \ \& \ A(t/x),$$

since in an empty domain with $A(t/x)$ false, the former but not the latter is true.

Despite these complications, free logics are often inclusive. Free logicians are generally romantics at heart, chafing at the slightest existential presumption and preferring the wild possibilities of emptiness to the neat regularities of classical restraint.

3 PROOF THEORY

Inclusive free logic lends itself readily to natural deduction. The formation rules are just those of classical predicate logic, but there must be an additional clause for the existence predicate '$E!$' if it is taken as primitive rather than defined in terms of the identity predicate. Either way, the quantifier rules for the standard bivalent system of inclusive free logic (for closed formulas) are as follows:

$\forall$**I**: Given a derivation of $A(t/x)$ from $E!t$, where t is new and does not occur in A, discharge $E!t$ and infer $\forall xA$.

$\forall$**E**: From $\forall xA$ and $E!t$ infer $A(t/x)$.

$\exists$**I**: From $A(t/x)$ and $E!t$ infer $\exists xA$.

$\exists$**E**: Given $\exists xA$ and a derivation of a formula B from $A(t/x)$ & $E!t$, where t is new and does not occur in either A or B, discharge $A(t/x)$ & $E!t$ and infer B from $\exists xA$.

($A(t/x)$ is the result of replacing all occurrences of x in A by t. If t does not occur in A, then $A(t/x)$ is just A.) Rules for the propositional operators are just as in classical logic. No special introduction or elimination rules are needed for '$E!$'. If the identity predicate is present, its elimination rule is the usual principle of substitutivity, but its introduction rule may be nonstandard. Classical $=$**I** permits the introduction (on no assumptions) of any statement of the form $t = t$. But where t is nondenoting $t = t$ may, on negative semantics (discussed below), be untrue. On such semantics, $=$**I** can introduce statements of the form $t = t$ only on the assumption $E!t$. (An alternative, but equivalently restricted $=$**I** rule introduces, on no assumptions, statements of the form $\forall xx = x$.) If, however, we require all self-identities to be true, regardless of whether or not the term denotes, then $=\mathbf{I}$ can have its classical form, allowing us to introduce formulas of the form $t = t$ on no assumptions.

The same logic may be axiomatized by adding the following schemas to the tautologies of classical propositional logic:

(A1) $A \rightarrow \forall xA$

(A2) $\forall x(A \rightarrow B) \rightarrow (\forall xA \rightarrow \forall xB)$

(A3) $\forall xA \rightarrow (E!t \rightarrow A(t/x))$

(A4) $\forall xE!x$

(A5) $\forall xA(x/t)$ if A is an axiom.

(Once again, all formulas are presumed closed.) Modus ponens is the only necessary rule of inference. (A1), (A2) and (A5) are also classical axioms. (A3) and (A4) are peculiar to free logic. (A3) is a restricted form the classical principle of specification:

(Spec) $\forall xA \rightarrow A(t/x)$.

(A4) stipulates that the quantifiers range only over existing things. Replacing both (A3) and (A4) with (Spec) yields classical predicate logic.

If the identity predicate is included, we also need:

(A6) $s = t \rightarrow (A \rightarrow A(t//s))$

(where $A(t//s)$ is the result of replacing one or more occurrences of s in A by t) and either

(A7) $\forall x\ x = x$,

if only those self-identity statements whose singular term denotes a member of the quantificational domain need be true, or

(A7$^+$) $t = t$,

if we want all self-identity statements to be true. (A1)–(A7) form the core of inclusive free logic with identity (for closed formulas). A noninclusive free logic may be obtained by adding to (A1)–(A7) the axiom $\exists xE!x$ or its equivalent $\exists xx = x$. Additional axioms or rules may be needed if we adopt special semantic conventions, as will be explained below. If 'E!' is defined in terms of the identity predicate, (A3) becomes redundant, since it then follows from (A6) together with the other axioms.

It is possible to develop free logic without either '=' or 'E!'. Such a logic retains (A1), (A2) and (A5), but instead of (A3) and (A4) has:

(A3′) $\forall y(\forall xA \rightarrow A(y/x))$

(A4′) $\forall x \forall y A \rightarrow \forall y \forall x A$

(A3′), like (A3), is a version of the law of specification that restricts specification to objects within the domain of the quantifiers, though unlike (A3), it is expressed without the use of '$E!$'. The quantifier permutation axiom (A4′) is redundant in the presence of the identity axioms but is independent of (A1),(A2),(A3′) and (A5).[6]

Among free logics whose formulas may contain free variables, those that require the variables to denote objects in the quantificational domain are known as **$\mathbf{E}^+$**-*logics*; those that do not are called **E**-*logics*. Free variables in E-logics function merely schematically, so that, for example, E-logics have the substitution rule:

$$A \vdash A(t/x)$$

while E^+-logics do not. Conversely, E^+-logics validate $E!x$, for any individual variable x, while E-logics do not.[7] In what follows, however, we shall generally presume that formulas are closed, so that this distinction is irrelevant.

4 SINGLE VS. DUAL DOMAINS

Free logics differ semantically on the question of whether singular terms that are not assigned objects in the quantificational domain (hereafter called *empty* terms) refer to objects not in the domain or to nothing at all. There are two main approaches, which I shall call *dual-domain* and *single-domain* semantics, respectively.

In a single-domain semantics *the* domain typically represents the class of existing things, and empty singular terms have no referents. The interpretation function for a single-domain semantics is, in other words, partial. In a dual-domain semantics the interpretation function is usually total, but singular terms may refer to objects outside the quantificational domain. These outlying objects are collected into a second or *outer* domain, in contrast to which the usual quantificational domain is described as *inner*. The inner domain typically represents the class of existing things; the outer, correlatively, includes nonexistents. Thus on a dual-domain semantics all singular terms refer, but not all refer to existing things; the term 'empty' is therefore not equivalent to 'nonreferring' or 'nondenoting', as it is in single-domain semantics. Some authors make the inner domain a subset of the outer, which thus contains both existent and nonexistent objects; others make the two disjoint.

Many authors call single-domain semantics 'Russellian' and dual-domain semantics 'Meinongian', but these designations are misleading. Although the single/dual contrast is reminiscent of their contrasting ontologies, neither philosopher invented the semantics bearing his name. Moreover, the term "Meinongian semantics" has lately become ambiguous. Some authors now use it specifically to describe semantics equipped with quantifiers over all the denizens of some rich Meinongian theory of objects.[8] But the outer domain of a dual-domain semantics need not (and usually does not) have its own quantifiers; and, though it may include incomplete and impossible objects *à la* Meinong, it may also, depending on the application, contain only specific classes of nonexistents — for example: merely possible objects (in alethic modal logics), objects existing only in the past or future (in tense logics), undefined quantities (in mathematics), error objects (in computer science), or nonexisting theoretical entities (in philosophy and natural science).

---

[6] Fine [1983].

[7] For a more complete treatment of free logics with free variables, see Troesltra and van Dalen [1988], vol. I, ch. 2, sec. 2.

Because the interpretation function of a dual-domain logic is typically total, there must be something in the outer domain for singular terms to denote. Therefore, the outer domain must, like the domains of classical logic, be nonempty. Both single- and dual-domain semantics validate (A1)-(A7).

In addition to the single-/dual-domain divide, free logics differ in how they assign truth-values to atomic formulas containing empty singular terms. There are three general approaches. Call any formula containing one or more empty singular terms *referentially deficient*. Then a semantics for free logic is *positive* if it allows some referentially deficient atomic formulas not of the form E!t to be true, *negative* if it requires all such formulas to be false, and *nonvalent* if it requires them all to be truth-valueless.[9] Dual-domain logics are usually positive, for there is seldom reason to have an outer domain of nonexistent objects unless they can satisfy predicates. Single-domain logics are usually negative or nonvalent, but may be positive as well.

5 POSITIVE SEMANTICS

In a positive semantics, referentially deficient atomic formulas may be either true or false. Our definition does not rule out their having other truth-values as well, or truth-value gaps; but additional truth-values add nothing of direct interest here, and truth-valuelessness will be discussed in the section on nonvalent semantics below. The positive semantics considered in this section are bivalent. In positive dual-domain free semantics (where empty singular terms are assigned members of the outer domain), predicates are interpreted over the union of the two domains, and the truth-values of atomic formulas (whether referentially deficient or not) are computed in the usual Tarskian fashion: an atomic formula is true if and only if the n-tuple of objects denoted by its singular terms, taken in order, is a member of the predicate's extension. In positive single-domain free logics (whose empty singular terms are assigned no object at all), referentially deficient atomic formulas require some unconventional treatment.

---

[8] See, for example, Paniczek [1998] and [2001], Jacquette [1996], Lambert [1983], p. 97, or Routley [1980].

[9] This classification is Lambert's [2003f], pp. 126-7. Others (e.g., Morscher & Simons [2001]) use 'neutral' instead of 'nonvalent', but Routley [1980], pp. 76 ff., has claimed that term to describe logics whose "existential" quantifiers do not express existence. The term 'neuter' is another alternative (e.g., in Lambert [1991], p. 9), but not a very happy one.

In order to preserve the full logic of identity, self-identities are taken to be true in positive free logics, regardless of whether or not the term denotes. Hence the axiom (A7), which implies only identities between nonempty terms, is replaced in positive logics with the stronger (A7$^+$), which asserts all self-identity statements.

Given bivalence, $s = t$ must be false when only one of s and t denotes. Hence any bivalent semantics, whether positive or negative, validates the formula:

$$(E!s \ \& \ \sim E!t) \rightarrow \ \sim s = t.$$

But this requires no new axioms, for it follows from (A6).

When s and t are both empty, we may stipulate that $s = t$ is always true, but this is overly restrictive for some applications, and positive free logics often allow identities between distinct empty terms to be either true or false. For examples, see the section on definite descriptions below.

The simplest positive free semantics is a bivalent dual-domain semantics. Let $\mathcal{L}$ be a first-order language without sentence letters, with '$\sim$', '&', '$\forall$', '$E!$' and '$=$' primitive and the other logical operators defined as usual. Then a dual-domain model for $\mathcal{L}$ may be defined as a triple $\langle \mathcal{D}_i, \mathcal{D}_o, I\rangle$, consisting of a possibly empty inner domain $\mathcal{D}_i$, a nonempty outer domain $\mathcal{D}_o$ such that $\mathcal{D}_i \subseteq \mathcal{D}_o$, and an interpretation function I such that for every individual constant t of $\mathcal{L}, I(t) \in \mathcal{D}_o$ and for every n-place predicate P of $\mathcal{L}, I(P) \subseteq \mathcal{D}_o^n$. Given a model $\mathcal{M} = \langle \mathcal{D}_i, \mathcal{D}_o, I\rangle$, truth-values are assigned by the valuation function $\mathcal{V}$ such that:

$\mathcal{V}(Pt_1 \ldots t_n) = T$ iff $\langle I(t_1), \ldots, I(t_n)\rangle \in I(P); = F$ otherwise.
$\mathcal{V}(s = t) = T$ iff $I(s) = I(t); = F$ otherwise.
$\mathcal{V}(E!t) = T$ iff $I(t) \in \mathcal{D}_i; = F$ otherwise.
$\mathcal{V}(\sim A) = T$ iff $\mathcal{V}(A) = F; = F$ otherwise.
$\mathcal{V}(A \ \& \ B) = T$ iff $\mathcal{V}(A) = T$ and $\mathcal{V}(B) = T; = F$ otherwise.
$\mathcal{V}(\forall xA) = T$ iff for any individual constant t not occurring in A and all
    $d \in \mathcal{D}_i, \mathcal{V}_{(t,d)}(A(t/x)) = T; = F$ otherwise.

(Here $\mathcal{V}_{(t,d)}$ is the valuation function on the model $\langle \mathcal{D}_i, \mathcal{D}_o, I^*\rangle$ such that I^* is just like I except that $I^*(t) = d$.) The logic defined by the natural deduction rules mentioned above together with the usual identity rules or by (A1)–(A6) and (A7$^+$) is sound and complete with respect to this semantics.[10]

The nonexistent objects of a dual-domain semantics often serve to generalize, systematize and simplify theory. In this respect they function like the ideal points at infinity of projective geometry. Dual-domain semantics are, however, open to the charge of ontological extravagance. Single-domain positive semantics avoid this objection. The literature offers two main varieties.[11] The first, introduced by van Fraassen [1966], bypasses the usual Tarski definition of truth for referentially deficient atomic formulas and assigns them truth-values directly by fiat. These truth-value assignments are, however, mere preliminaries to the construction of supervaluations and so will be discussed in the section on supervaluations below.

---

[10] The proof can be found in Leblanc and Thomason [1968].

[11] There are others, less general in application. See, for example, the discussion of Scott's single-domain positive free semantics for Quine's virtual theory of classes in the section on free

The second single-domain form of positive semantics is the "proto-semantics" recently proposed by G. Aldo Antonelli [2000]. A proto-semantics provides two ways of determining the truth-values of atomic formulas: a Tarksi-style determination for formulas whose terms all refer and a stipulated truth-value assignment for referentially deficient formulas. A model in proto-semantics is a triple $\langle \mathcal{D}, \rho, I\rangle$, where $\mathcal{D}$ is a possibly empty domain of objects (all of which are regarded as existing); ρ is a partial reference function such that for any individual constant t, $\rho(t)$ is either undefined or a member of $\mathcal{D}$; and I is an interpretation function that assigns to each n-tuple ($n \geq 1$) of individual constants a proto-interpretation. A proto-interpretation is an assignment of a signed extension (that is, a pair $\langle \mathcal{E}, \mathcal{S}\rangle$ where E is an ordinary extension (a subset of $\mathcal{D}^n$) and $\mathcal{S}$ is one of the two signs $+$ or -) to each n-place predicate. The signs ($+$ for T and - for F) encode the truth-values that atomic formulas take in case of reference failure. Let I be an interpretation with reference function ρ, π be the proto-interpretation that I assigns to the n-tuple $\langle t_1, \ldots, t_n\rangle$ of individual constants, and P be an n-place predicate such that $\pi(P) = \langle \mathcal{E}, \mathcal{S}\rangle$. Then if at least one of $t_1, \ldots, t_n$ is empty, I makes the atomic formula $Pt_1 \ldots t_n$ true if $\mathcal{S} = +$ and false if $\mathcal{S} = \text{-}$. But if $t_1, \ldots, t_n$ all refer, then truth is determined conventionally; that is, I makes $Pt_1 \ldots t_n$ true if $\langle t_1, \ldots, t_n\rangle \in \mathcal{E}$, false otherwise.

The proto-interpretations assigned to n-tuples of terms by a given interpretation are subject to two constraints: (1) all must all agree on the extensions they assign to predicates; they therefore differ from one another, if at all, only in their assignments of signs, and (2) to preserve substitutivity of identicals, if individual constants s and t are assigned the same proto-interpretation, then any two n-tuples of individual constants differing only in that one has an occurrence of s where the other has t must be assigned the same proto-interpretation.

Antonelli uses a substitutional interpretation of quantifiers, which requires that for each member d of $\mathcal{D}$, there be an individual constant c_d that names it (i.e., $\rho(c_d) = d$). Given a model $\mathcal{M} = \langle \mathcal{D}, \rho, I\rangle$, a valuation function $\mathcal{V}$ assigns truth-values as follows:

> Let $I(\langle t_1, \ldots, t_n\rangle) = \pi$ and $\pi(P) = \langle \mathcal{E}, \mathcal{S}\rangle$. Then if $\rho(t_i)$ is undefined for some $i, 1 \leq i \leq n$, then $\mathcal{V}(Pt_1 \ldots t_n) = T$ if $S = +$ and $\mathcal{V}(Pt_1 \ldots t_n) = F$ if $S = \text{-}$. But if $\rho(t_i)$ is defined for all $i, 1 \leq i \leq n$, then $\mathcal{V}(Pt_1 \ldots t_n) = T$ if $\langle t_1, \ldots, t_n\rangle \in E$ and $\mathcal{V}(Pt_1 \ldots t_n) = F$ if $\langle t_1, \ldots, t_n\rangle \notin \mathcal{E}$.
>
> $\mathcal{V}(s = t) = T$ iff either both $\rho(s)$ and $\rho(t)$ are defined and $\rho(s) = \rho(t)$, or both $\rho(s)$ and $\rho(t)$ are undefined and $I(s) = I(t)$; $= F$ otherwise.

---

set theory below. There is also the truth-value semantics of LeBlanc [1971], but since it dispenses with domains altogether and uses substitutional quantifiers, it is neither single-domain nor free in the sense defined above.

$\mathcal{V}(E!t) = T$ iff $\rho(t)$ is defined; $= F$ otherwise.

$\mathcal{V}({\sim}A) = T$ iff $\mathcal{V}(A) = F$; $= F$ otherwise.

$\mathcal{V}(A \mathbin{\&} B) = T$ iff $\mathcal{V}(A) = T$ and $\mathcal{V}(B) = T$; $= F$ otherwise.

$\mathcal{V}(\forall xA) = T$ iff for all $d \in \mathcal{D}, \mathcal{V}(A(c_d/x)) = T$; $= F$ otherwise.

Axioms (A1)–(A6) and ($A7^+$) are all valid on proto-semantics, and Antonelli offers soundness, completeness, compactness and interpolation theorems. Proto-semantics thus delivers a positive free logic while avoiding the ontological profligacy of outer domains — but at the cost of increased artifice and semantic complexity.

It is subject, moreover, to a difficulty that infects all free positive semantics: when applied, all such semantics force us (whether by interpretation of predicates over outer domains, by fiat or by proto-interpretations) to decide the truth-values of many referentially deficient atomic formulas more or less arbitrarily. Is the statement '$1/0 = 2/0$', for example, true? How about '$2/0 > 1/0$' or '$1/0 > 1$'? It is tempting to answer "no" to all such questions; but, given bivalence, that way lies negative semantics. Answering "yes" to any of them yields a positive semantics, but then we face the task of constructing an intelligible and consistent yet varied set of answers to many such questions. Depending on the application, that may not be worth the trouble.

6 NEGATIVE SEMANTICS

If it isn't, we may prefer a negative semantics. Negative semantics answer all such questions with a simple, appealing and uniform stipulation: all referentially deficient atomic formulas are false. Consequently, in any negative free logic containing some version of '$E!$' the following schema is valid:

(A−) $Pt_1 \ldots t_n \rightarrow (E!t_1 \& \ldots \& E!t_n)$, where P is a primitive n-place predicate.

For completeness' sake, negative free logics must include (A−) or its equivalent as an additional axiom or rule.

Since the identity predicate, if taken as primitive, is an instance of P,

$$s = t \rightarrow (E!s \mathbin{\&} E!t)$$

and

$$t = t \rightarrow E!t$$

are instances of (A−). The latter formula together with (A2), (A5) and (A7) implies (A4), so that in a negative free logic with identity (A−) may replace (A4). Moreover, since the converse formula:

$$E!t \rightarrow t = t$$

follows from (A3) and (A7), negative free logics validate the equivalence:

$$t = t \leftrightarrow E!t.$$

$E!t$ may therefore be defined simply as $t = t$.

Because they make all referentially deficient atomic formulas false, negative free logics validate the principle of *indiscernibility of nonexistents*:

$$(\sim E!s \;\&\; \sim E!t) \rightarrow (A \rightarrow A(t//s)).$$

(Here, as in (A6), $A(t//s)$ is the result of replacing one or more occurrences of s in A by t.)

But this principle need not be posited as a separate axiom, for all its instances follow from (A−).[12]

A negative single-domain model for a language $\mathcal{L}$ is a pair $\langle \mathcal{D}, I \rangle$, consisting of a possibly empty domain $\mathcal{D}$ and a partial interpretation function I such that for each individual constant t of $\mathcal{L}$ either $I(t) \in \mathcal{D}$ or $\mathcal{I}(t)$ is undefined, and for every n-place predicate P of $\mathcal{L}, I(P) \subseteq \mathcal{D}^n$. Given a model $\mathcal{M} = \langle \mathcal{D}, \mathcal{I} \rangle$, truth-values are assigned by a valuation function $\mathcal{V}$, defined as follows:

> $\mathcal{V}(Pt_1 \ldots t_n) = T$ iff $I(t_1), \ldots, I(t_n)$ are all defined and $\langle I(t_1), \ldots, I(t_n) \rangle \in I(P)$; $= F$ otherwise.
>
> $\mathcal{V}(s = t) = T$ iff $I(s)$ and $I(t)$ are both defined and $I(s) = I(t)$; $= F$ otherwise.
>
> $\mathcal{V}(E!t) = T$ iff $I(t)$ is defined; $= F$ otherwise.
>
> $\mathcal{V}(\sim A) = T$ iff $V(A) = F$; $= F$ otherwise.
>
> $\mathcal{V}(A \;\&\; B) = T$ iff $\mathcal{V}(A) = T$ and $V(B) = T$; $= F$ otherwise.
>
> $\mathcal{V}(\forall x A) = T$ iff for any individual constant t not occurring in A and all $d \in \mathcal{D}, \mathcal{V}_{(t,d)}(A(t/x)) = T$; $= F$ otherwise.

(Here $\mathcal{V}_{(t,d)}$ is the valuation function on the model $\langle \mathcal{D}, I^* \rangle$ such that I^* is just like I except that $I^*(t) = d$.) The logic defined by (A1)–(A7) and (A−) is sound and complete with respect to this semantics.

In negative free logics referential deficiency guarantees the falsity only of *atomic* formulas. Logical truth is thus severed from logical form. For let F be a primitive one-place predicate; then although $Ft \rightarrow E!t$ is logically true, substitution instances — $\sim Ft \rightarrow E!t$, for example — in which F is replaced by a complex predicate are contingent.

Tyler Burge has defended (A−) as a "methodological condition on investigations of predication in natural language," whose application he describes as follows:

> count an expression an atomic [i.e., primitive] predicate in natural language only if one is prepared to count simple singular sentences containing it untrue whenever they also contain nondenoting terms.[13]

---

[12] This is true, at least, for negative logics without function symbols. In a negative logic with function symbols, proof of the indiscernibility of nonexistents also requires the strictness axiom, discussed in the section on definedness logics and programming languages below.

[13] Burge [1974] in Lambert [1991], p. 193.

If, for example, we account the sentence ‘Chrissy loves Santa Claus’ true, then, assuming Santa does not exist, we must on Burge’s view regard the predicate ‘loves’ as non-primitive. But if natural languages have any primitive predicates, why shouldn’t ‘loves’ be among them? It is doubtful, moreover, that the primitive/non-primitive distinction can successfully be made in natural languages, for it is closely bound up with the now largely discredited doctrine of logical atomism. (A−) cannot, in any case, be *merely* a methodological condition; for if only those models are admissible that make all referentially deficient atomic formulas false, then all instances of (A−) are full-blooded logical truths.

A related anomaly is the dependence of the truth-values of referentially deficient statements upon our choice of primitives. Because negative free logic makes referentially deficient atomic formulas false, it makes their negations true. Now consider a negative free logic interpreted over the domain of natural numbers that takes as primitive a one-place predicate O meaning “is odd” and defines “is even” by this schema:

$$Et =_{df} {\sim} Ot$$

For any nondenoting term t, Ot is false in this theory; hence Et is true. Take a predicate meaning ‘is even’ as primitive instead and the truth-values are reversed. But why should the truth-value of any such statement hinge on our choice of primitives? Worse, why should our choice of primitives determine whether the statements of the form ‘If t is even, then t exists’ are logical truths?

If we want reference failure to make atomic statements false but choice of primitives seems too evanescent a matter on which to hang logical truth, we could make *any* formula containing an empty term false. A semantics implementing this idea — call it a *supernegative* semantics — would validate no formulas containing singular terms but would yield a logic that is ordinary in other respects. ($Ft \vee {\sim} Ft$, for example, would not be valid on a supernegative semantics, though $\forall x(Fx \vee {\sim} Fx)$ would be.) Such a semantics would be reminiscent of Russell’s logic of names, which contextually analyzes names into definite descriptions, on a wide-scope reading of the descriptions. This reading would likewise render all schemata containing names invalid, since failure of the descriptions’ existence or uniqueness conditions could always make these schemata false (see the discussion of Russell’s theory of descriptions below). But no one, so far as I know, has advocated supernegative semantics. Presumably this is because a free logic must not merely block all inferences that are invalid if our dubious terms turn out not to refer; it should also maximize preservation of inferences that are valid if they do refer. Supernegative free logics would preserve too few.

But even ordinary negative free logics preserve too few. They invalidate, for example, the formula $t = t$. Since formulas of this form are true when their singular terms denote and since it is a matter of choice how they come out otherwise, why not make them true either way? Then if we are reasoning with a dubious term t that in fact denotes we won’t be blocked from inferring the logical truth $t = t$.

Moreover, while negative free logic preserves most other *logical* principles containing singular terms, it falsifies many *nonlogical* principles that depend on the semantic interrelations of primitive predicates. In negative free number theory, for example, all referentially deficient instances of the law of trichotomy:

$$s = t \lor s < t \lor t < s$$

are false, assuming that '<' and '=' are primitive. If we won't tolerate falsifications of familiar laws in logic, why tolerate them in number theory? The number theorist's singular terms (which include, for example, functional expressions for limits that fail to denote if the limit does not exist) all *aim* to denote numbers that obey the law of trichotomy. Why not arrange the semantics so that this law remains true whether they achieve that aim or not? (We could, for example, stipulate that all identities between nondenoting terms are true and that $s < t$ is true whenever s denotes and t does not.) That, however, would make our semantics positive.

At the very least, why not arrange things so that no instance of such desirable principles as the law of trichotomy is ever false?

7 NONVALENT SEMANTICS

Nonvalent semantics avoid falsifying desirable principles by making all referentially deficient atomic formulas (except those of the form E!t) truth-valueless. Truth-value gaps seem appropriate in a single-domain free semantics, for partial truth-value assignments are a natural result of partial denotation assignments.

But nonvalence sets before us a bewildering multiplicity of semantic choices. For example:

- When one disjunct is true and the other truth-valueless, should a disjunction be true or truth-valueless?
- When one disjunct is false and the other truth-valueless, should a disjunction be false or truth-valueless?
- When A is truth-valueless for each object in the domain, should $\exists xA$ be false or truth-valueless?
- When A is either false or truth-valueless for each object in the domain, should $\exists xA$ be false or truth-valueless?

Nonvalent semantics implementing each of these choices in various combinations have appeared in the literature.[14]

Of the various nonvalent semantics, one of the simplest and most radical is the No Input No Output (NINO) semantics advocated by Scott Lehman ([1994] and [2000]), who attributes the basic idea to Frege. On a NINO semantics, which has only one domain, predicates and logical operators represent partial functions that fail to yield a value if any of the semantic values ("input") to which they are applied is not defined. Thus referentially deficient atomic formulas (including identity statements) are truth-valueless. (As usual, '$E!$', if included in the language, is an exception, since $E!t$ is true if t is empty.) Disjunctions with a truth-valueless disjunct also lack necessary input and are likewise truth-valueless. But all quantified statements are either true or false, even if they contain empty terms. This is because Lehman regards quantifiers as naming functions from partial functions (the designata of open formulas) to truth-values. Since these partial functions exist even when their output is undefined for all inputs, the functions represented by quantifiers always have input and hence always yield a truth-value as output.

---

[14] For a summary, see Lehman [1994], p. 328.

NINO semantics is akin to the supernegative semantics discussed above, but rather than making all referentially deficient formulas false, it makes them all (except for quantified formulas) truth-valueless. Thus it preserves more of the structure of logic, for while on a supernegative semantics such classically valid expressions as $t = t$ and $\sim(Ft \;\&\; \sim Ft)$ are false if t is empty, and hence are in no sense valid, on a NINO semantics they are merely truth-valueless and hence remain *weakly valid* — i.e., not false on any model. Moreover, as on a supernegative semantics, the formulas obtained from them by universal generalization are *strongly valid* — i.e., true on all models.

NINO semantics makes some classical principles — e.g., $\exists xA \vee \sim \exists xA$ — strongly valid, most weakly valid, and some — e.g., $\exists x\; x = t$ or $\exists xA \vee \exists x \sim A$ — not even weakly valid. Lehman thinks weak validity is all we need and argues for a logic based on weak rather than strong validity.

But disconcerting consequences follow. For one thing, if logical truth is weak rather than strong validity, logically true statements need not even be true.[15] Furthermore, a NINO logic based on weak validity is in some respects too strong. The schema:

$$A(x/t) \rightarrow \exists xA,$$

for example, which free logicians reject, is weakly valid on NINO semantics.

If the NINO strategy thus seems unpromising, we may drop it, retain strong validity as the logical standard, and design a nonvalent semantics to make as many desirable laws as possible strongly valid. We can, for example, make all self-identities true and account disjunctions true whenever a disjunct is true. But the resulting logic will still be discouragingly weak, failing to validate such standard principles as $\sim(Ft \;\&\; \sim Ft)$.

The problem is that truth-value gaps at the atomic level continue, despite such semantic tinkering, to be transmitted to more complex formulas in ways that invalidate desirable logical laws. To create stronger nonvalent logics (assuming a strong conception of validity), we must more effectively limit this upward transmission of truth-valuelessness, enabling logical laws to remain true even when their components lack truth-value.

---

[15] Lehman is aware of this consequence and accepts it. See Lehman [2001], p. 153.

8 SUPERVALUATIONS

Supervalutions do just that. The idea is that when reference failure deprives atomic formulas of truth-value, complex formulas built from them should count as true (or false) if all ways of "filling in" the missing truth-values agree in making them so.

A supervaluation begins with a single-domain model M in which atomic formulas have truth-value if and only if they are not referentially deficient. In Bas van Fraassen's original [1966] conception, we then consider various assignments of one of the values T or F to each referentially deficient atomic formula. In order to validate the laws of identity and quantification, these assignments must meet certain formal constraints, the details of which need not concern us here.[16] Call the original model $\mathcal{M}$ augmented by such a truth-value assignment a *bivalent* $\mathcal{M}$-*model*. Then, for example, if the term t is empty in $\mathcal{M}$, the atomic formula Ft will be truth-valueless on $\mathcal{M}$ but either true or false on each bivalent $\mathcal{M}$-model; hence the complex formula $Ft \vee \sim Ft$ will be true on all bivalent $\mathcal{M}$-models.

Now the *supervaluation* of $\mathcal{M}$ is a partial assignment of truth-values to formulas that makes a formula true if all bivalent $\mathcal{M}$-models make it true, false if they all make it false, and truth-valueless if they fail to agree. On the supervaluation of $\mathcal{M}$, for example, Ft is truth-valueless but $Ft \vee \sim Ft$ is true. A formula is *logically* true on a supervaluational semantics if and only if it is true on all supervaluations. In this way, supervaluations block the transmission of truth-value gaps from atomic formulas to logical laws that contain them.

They do this so well, in fact, that (depending on the constraints we place on the truth-value assignments augmenting $\mathcal{M}$ and how empty domains are handled) they may validate all *classical* theorems.[17] Van Fraassen shows, however, how to construct bivalent $\mathcal{M}$-models that validate only the theorems of positive free logic (those implied by (A1)–(A6) and ($\mathrm{A7}^{+}$)).

Van Fraassen's method of assigning truth values by fiat strikes some as contrived, but Ermanno Bencivenga has developed a more natural alternative.[18] Like van Fraassen, Bencivenga begins with single domain model $\mathcal{M}$ that is nonvalent, except it makes all identities between denoting and nondenoting terms false, rather than undefined. He then assigns truth-values to complex referentially deficient formulas depending upon what their truth-values *would be* if their singular terms all denoted. More specifically, he considers all *complete supermodels* of $\mathcal{M}$ (i.e., bivalent single-domain models whose domain is a superset $\mathcal{S}$ of the domain of $\mathcal{M}$, on which all singular terms denote members and each n-place predicate is true or false of each n-tuple of memberss of $\mathcal{S}$. Freedom is guaranteed by allowing the quantifiers of these supermodels to range only over the original domain of $\mathcal{M}$. The supervaluation of $\mathcal{M}$ is then defined as before: a formula is true on this supervaluation if all complete supermodels of $\mathcal{M}$ make it true, false if they all make it false, and undefined if they fail to agree. This semantics, like van Fraassen's validates all the theorems of positive free logic.[19]

---

[16] For an account of the details, see Bencivenga [1985] pp. 402-3.

[17] See Bencivenga [1985] for details.

[18] Bencivenga [1981] and [1985]. See also Morscher and Simons [2001], pp. 14–18.

There is, however, much not to like about supervaluations, no matter how they are constructed. For one thing, they embody a strange conception — Bencivenga [1985] calls it the "counterfactual theory" — of truth. Consider, for example, the sentence 'This car is a Ford' in a context in which no car is indicated. Since 'this car' is empty in that context, it is reasonable to regard the sentence as truth-valueless. But then in the same context the disjunction 'Either this car is a Ford or this car is not a Ford' seems equally truth-valueless. The notion that this disjunction is true because it would be true if 'this car' denoted (Bencivenga) — or if 'this car is a Ford' had truth-value (van Fraassen) — is dubious at best. Indeed, 'the counterfactual theory of truth' is, arguably, an oxymoron.

There are also logical difficulties. For one thing, substitutivity *salva veritate* fails even at the propositional level under supervaluations. Where A is truth-valueless, $A \rightarrow A$ is logically true on a supervaluational semantics. But if we substitute a new truth-valueless formula atomic B for the second occurrence of A, the result, $A \rightarrow B$, is not true but truth-valueless. (This sort of failure is peculiar to supervaluations; it can be avoided, for example, by handling referential deficiency via a third truth-value in a truth-functional semantics.)

Supervaluational consequence, moreover, is too strong. Though the formula $Ft \rightarrow E!t$ is not logically valid in either van Fraassen's or Bencivenga's semantics, both semantics make $E!t$ true on every supervaluation on which Ft is true, thus validating the *inference* from Ft to $E!t$. Therefore, although positive free logic is sound and weakly complete on these semantics, it is not strongly complete.[20] In fact, by defining a translation between supervaluational and second-order formulas, Peter Woodruff [1984] has shown that supervaluational predicate logics are in effect second order. It follows, then, not only that compactness fails for supervaluational semantics but also that supervaluational consequence is not axiomatizable by any extension of free logic.[21]

9 WHICH SEMANTICS?

Though nonvalent and supervaluational semantics yield disfigured logics, this is not sufficient reason to dismiss them. For some applications, nothing elegant will do. It does, however, establish a presumption in favor of either positive or negative semantics.

But positive semantics, as I noted above, can be unwieldy in application, forcing us to decide case by case the truth-values of referentially deficient atomic formulas.

---

[19] Morscher & Simons [2001], pp. 17–18.

[20] A logic is *strongly complete* if for any finite set S of premises and conclusion C, if C is semantically implied by S, then C is deducible from S in the logic; it is *weakly complete* if every semantically valid formula is a theorem of the logic.

[21] *Compactness* holds if for any set S of formulas, S semantically implies formula C if and only if some finite subset of S semantically implies C.

Many of these cases may be "don't cares" that we would rather ignore — and can ignore in negative or nonvalent semantics. One way to retain the best features of both positive and negative logics is to adopt the "default rule" that referentially deficient atomic formulas are false *unless otherwise specified.* Then, though the resulting semantics is positive, allowing us to make referentially deficient formulas true whenever we please, there is, as in negative semantics, no need to bother about the "don't cares." Moreover, because it is positive, this semantics also avoids both the entanglement of logical truth with primitiveness that mars negative logics and their nonstandard identity theory.

But no semantics is best for all applications. Positive semantics offer the greatest flexibility and so tend to excel in applications close to natural language. But for more formal tasks, the simplifying regimentation imposed by negative or nonvalent logics may fill the bill. And where unavoidable truth-value gaps eviscerate logic, supervaluations can ameliorate the damage.

10 DEFINITE DESCRIPTIONS

Free logic has been most extensively and instructively applied in the theory of definite descriptions. A definite description is, as Lord Russell intones, "a phrase of the form the-so-and-so in the singular" — e.g.: 'the smallest positive integer', 'the current American President', 'the element with atomic number 23'. To formalize definite descriptions, Russell borrowed Peano's variable-binding description operator 'ι' (usually written inverted). When prefixed to a variable x and a formula A in which x is free, 'ι' forms a singular term, ιxA, meaning "the x such that A." In *Principia*, definite descriptions are introduced for most contexts by the definition:

$$B(\iota xA/y) =_{df} \exists y(\forall x(A \leftrightarrow x = y) \;\&\; B), x \text{ free in } A,\ y \text{ free in } B.^{22}$$

That is, the A is B iff there is one and only one A and it is B. This definition produces scope ambiguities unless augmented by some device to mark the scope of the description. But its chief drawback from the point of view of free logic is that it builds an existence claim into every use of a definite description. That is, it implies:

$$B(\iota xA/y) \rightarrow \exists xA.$$

It is, however, often useful to reason with definite descriptions, as with proper names, even when we are not sure that they describe one and only one existing thing. Consider, for example, 'the person I will marry'.

Russell introduces applications of the predicate '$E!$' to definite descriptions by contextual definition, as follows:

$$E!\iota xA =_{df} \exists!xA, x \text{ free in } A.$$

---

[22] This is proposition 14.01 of *Principia.*

The uniqueness quantifier '$\exists!$' is in turn defined as:

$$\exists! xA =_{df} \exists y \forall x(A \leftrightarrow x = y), x \text{ free in } A,$$

a standard abbreviation that I shall employ again below.

In free description theories, by contrast, '$E!$' is often a primitive predicate and definite descriptions are genuine singular terms; neither is to be analyzed away by contextual definition. Set abstraction terms, for example, or such mathematical expressions as:

$$\lim_{x \to \infty} f(x)$$

may thus be singular terms, not contextually defined abbreviations, and yet they may still be empty. This is an advantage, for contextual definition obscures logical form.

Semantically, singular terms formed by the operator 'ι' function as follows: for any object d in the quantificational domain, ιxA denotes d iff d is the one and only object in the domain that satisfies A. Thus ιxA is empty if there exists more than one object satisfying A, or none. The denotation conditions for ιxA can be expressed in a free object language as identity conditions, by means of the following schema:

$$\text{(LL)} \quad \forall y(y = \iota xA \leftrightarrow \forall x(A \leftrightarrow x = y)), x \text{ free in } A,$$

which is known in honor of its presumed discoverer, Karel Lambert, as Lambert's Law.[23] All theories of definite descriptions agree on its validity.

It might seem that (LL) could be simplified and at the same time generalized to:

$$t = \iota xA \leftrightarrow \forall x(A \leftrightarrow x = t),$$

which aims to give the conditions under which an identity between *any* term (empty or not) and a definite description is true. But together with (A7^+), which is valid in positive logics, this schema implies that there exists no more than one object. For from the instance:

$$\iota x(x = x) = \iota x(x = x) \leftrightarrow \forall x(x = x \leftrightarrow x = \iota x(x = x)),$$

it follows by (A7^+) that:

$$\forall x(x = x \leftrightarrow x = \iota x(x = x)),$$

and hence that:

$$\forall x x = \iota x(x = x).^{24}$$

---

[23] Quine [1997], p. 104, says that Peano used this principle an axiom, but gives no reference. If Quine is right, then Peano was the discoverer and the law should bear his name. I have been unable to find any mention of such an axiom, however, in those of Peano's works that are readily available to me.

Nor can (LL) be replaced by:

$$t = \iota x A \leftrightarrow \forall x((A \rightarrow x = t) \ \& \ A(t/x)),$$

which Jaakko Hintikka [1959] once proposed as a principle of free description theory. For, putting $\iota x A$ for t, this schema together with (A7^{+}) implies unrestrictedly that the A is an A — in symbols:

$$A(\iota x A/x).$$

While superficially plausible, this schema is inconsistent, for when we substitute $(Fx \ \& \ {\sim}Fx)$ for A it immediately yields the contradiction:

$$F\iota x(Fx \ \& \ {\sim}Fx) \ \& \ {\sim}F\iota x(Fx \ \& \ {\sim}Fx),$$

which Lambert calls the *description paradox*.

Adding (LL) to the free logic defined by (A1)–(A7) gives the minimal free definite description theory MFD. Some of its most important theorems are:

$$\begin{array}{l} E!\iota x A \rightarrow A(\iota x A/x) \\ E!\iota x A \leftrightarrow \exists! x A \\ \forall x(A \leftrightarrow B) \rightarrow \forall y(y = \iota x A \leftrightarrow y = \iota x B) \\ (E!\iota x A \ \& \ B(\iota x A/x)) \leftrightarrow \exists y \forall x((A \leftrightarrow x = y) \ \& \ B(y/x)), \end{array}$$

where x is free in both A and B.

But given free logic's restriction on specification, the variable y in (LL) is instantiable only by nonempty singular terms. MFD therefore fails to answer the question, "When are identity statements between empty terms true?" There are many possible answers, limited by the extremes "always" and "never." "Never" yields a negative free theory of definite descriptions, characterized by the theorems:

$$\begin{array}{l} ({\sim}E!s \vee {\sim}E!t) \rightarrow \ {\sim}s = t \\ t = \iota x A \rightarrow A(t/x) \end{array}$$

and the Russell equivalence:

$$B(\iota x A/y) \leftrightarrow \exists y(\forall x(A \leftrightarrow x = y) \ \& \ B),$$

where B is restricted to atomic formulas — which restriction would in Russell's theory give description operators smallest possible scope. This negative free description theory may be axiomatized by adding (LL) to the negative free logic

---

[24] Nor is this schema appropriate for negative free logics, which have the weaker (A7) in place of (A7^{+}). For note that for any term t, the instance:

$t = \iota x(x = t) \leftrightarrow \forall x(x = t \leftrightarrow x = t)$

implies

$t = \iota x(x = t)$,

which in negative free logics implies $E!t$. This makes the logic classical.

characterized by (A1)–(A3), (A5)–(A7) and (A-). It is, in fact, the closest free cousin of Russell's theory of definite descriptions, though its singular terms lack existential import in nonatomic contexts and are not to be analyzed away by contextual definition.[25]

Allowing identity statements between empty terms *sometimes* to be true yields one or another positive theory — the simplest being the one on which they are *always* true. This theory, which Lambert has dubbed FD2, can be axiomatized by adding:

(FD2) $(\sim E!s \ \& \ \sim E!t) \rightarrow s = t.$

to (LL) and (A1)–(A7). Important theorems of FD2 include:

$$\begin{array}{l} (\sim E!t \ \& \ A(t/x)) \rightarrow A(\iota x A/x) \\ \forall x(A \leftrightarrow x = t) \rightarrow A(\iota x A/x) \\ \iota x(x = t) = t \\ \sim \exists! x A \rightarrow \iota x A = \iota x(x \neq x), \end{array}$$

where x is free in A and B.

In making all nonexistents identical, FD2, like negative free logic, makes them all indiscernible. Moreover, it approximates Frege's plan of choosing some arbitrary object to serve as the conventional denotation for all empty singular terms. (But since Frege's logic was not free, his implementation of the idea involved a problematic identification of nonexistent objects with an existent one; FD2 improves on Frege's idea in this respect.) FD2 can be modeled in a dual-domain semantics with only one object in the outer domain.

Like negative free description theories, FD2 precludes there being two empty descriptions $\iota x A$ and $\iota x B$ such that $F\iota x A$ and $F\iota x B$ differ in truth-value. For applications close to natural language however, we may want them to differ. The following statement, for instance, is true:

> Le Verrier sought the planet causing perturbations in Mercury's orbit,

while this one:

> Le Verrier sought the fountain of youth

is false — though each apparently applies the same predicate phrase 'Le Verrier sought' to an empty description.[26]

Thus while FD2 is handy for use in regimented theories that employ descriptions, more natural applications may require a positive free description theory on which identities between empty terms may be false. But then exactly which identities between empty terms are to be false and which true? Answers to this question are, as we have just seen, closely bound up with answers to another: how do nonlogical predicates interact with empty definite descriptions?

---

[25] See Burge [1974] for additional details. See also Lambert [2001h].

[26] See note 2.

Syntactic investigations of these questions have revealed a hierarchy of positive free description theories between the minimal theory, MFD, which imposes no restraints on the identity conditions of nonexistents and the maximal positive one, FD2, which makes all nonexistents identical. The theorems of FD2 listed above, for example, are pairwise nonequivalent, and each follows in MFD from the axiom (FD2). Hence each, if added to MFD, defines a logic stronger than MFD but weaker than FD2. Each therefore gives a partial answer to one or both of the questions just mentioned. Many other such theories are known, but the structure of the hierarchy is not fully known.[27]

These syntactic observations, however, provide little insight into the differing options as regards predication and identification. For that we must turn to semantics.

One plausible policy for making identity statements between empty terms sometimes true and sometimes false is to make $\iota xA = \iota xB$ true (whether ιxA and ιxB exist or not) if A and B have are *coextensive* (that is, if for all members d of the domain of quantification, d satisfies A if and only if d satisfies B). This may be formalized by adding the axiom:

$$\forall x(A \leftrightarrow B) \rightarrow \iota xA = \iota xB$$

to MFD. Lambert [2001a] calls the resulting system FDExt. Like FD2, FDExt implies:

$$\sim\exists xA \rightarrow \iota xA = \iota x(x \neq x),$$

but unlike FD2, it does not imply:

$$\sim\exists!xA \rightarrow \iota xA = \iota x(x \neq x).$$

Richard Grandy [1972] has proposed a related but weaker theory on which $\iota xA = \iota xB$ is true if A and B are *co-comprehensive* (that is, if for all members d of the union of the inner and outer domains, d satisfies A if and only if d satisfies B). This is most perspicuously expressed, given a dual-domain semantics, by introducing a universal quantifier that ranges over the union of the two domains. Using the generalized product symbol 'Π' for this quantifier, we may add the axiom:

$$\Pi x(A \leftrightarrow B) \rightarrow \iota xA = \iota xB$$

to MFD to obtain Grandy's theory, which Lambert [2001a] calls FDComp. FDComp can also be formulated without outer domain quantifiers by adopting this rule:

> Given a proof of $A(t/x) \leftrightarrow B(t/x)$ that uses no assumption containing t, infer $\iota xA = \iota xB$.

Both FDExt and FDComp lie within the hierarchy of systems that are intermediate in strength between MFD and FD2. But they are only two of many. The variety of positive free description theories seems inexhaustible.

---

[27] See Lambert [2001a] and [2003c] for further discussion.

11 SET THEORY

Free logic sometimes permits a crisp separation of those axioms of a theory that deal with the "essence" of their objects from those that deal with their existence. Nowhere is this more evident than in free set theory.

Sets are defined by their membership conditions, expressed as an open formula in one free variable. Thus where only x is free in A, the set defined by A is denoted by the set abstract notation $\{x \mid A\}$. In languages containing both the definite description operator 'ι' and the membership predicate '$\in$', set abstract notation can be defined as follows:

$$\{x \mid A\} =_{df} \iota z \forall x(x \in z \leftrightarrow A).$$

This definition elegantly links free set theory with free description theory, for given (LL) it implies:

$$E!\{x \mid A\} \rightarrow \forall y(y \in \{x \mid A\} \leftrightarrow A(y/x)),$$

a restriction of the "naïve" comprehension schema:

(Comp) $\forall y(y \in \{x \mid A\} \leftrightarrow A(y/x))$

to existing sets.

In classical logic, (Comp) is inconsistent, since by specification the instance

$$\forall y(y \in \{x \mid x \notin x\} \leftrightarrow y \notin y)$$

yields Russell's paradox:

$$\{x \mid x \notin x\} \in \{x \mid x \notin x\} \leftrightarrow \{x \mid x \notin x\} \notin \{x \mid x \notin x\}.$$

In free logic, however, the specification is invalid without the additional assumption

$$E!\{x \mid x \notin x\},$$

and this is not available. Thus in free logics (Comp) does not generate antinomies, but instead merely yields nonexistence proofs for paradoxical sets.

Unrestricted (Comp) can be obtained in positive free definite description theories by adding the axiom:

$$\sim E!\{x \mid A\} \rightarrow \forall y(y \in \{x \mid A\} \leftrightarrow A(y/x))$$

to MFD. But this axiom, which gives membership conditions for *non*existent sets, is generally not consistent with those positive free description theories stronger than MFD that impose their own differently motivated constraints on satisfaction of predicates by nonexistent objects. FD2 is an obvious case in point.

It is not necessary to couch free set theory within free description theory, however, and for the remainder of this section I will set free description theory aside, treating the set abstraction operator as primitive.

Together, (Comp) and the extensionality schema:

(Ext) $\forall z(z \in s \leftrightarrow z \in t) \rightarrow s = t$

characterize the "essence" of all sets, whether existent or not — (Comp) giving their membership conditions and (Ext) their identity conditions. These two principles, neither of which implies the existence of anything, therefore jointly constitute the core of positive free set theories — which differ from one another only in regard to which sets they assume to exist or not to exist. Such a neat bifurcation between axioms of essence and axioms of existence is impossible in classical logic, where comprehension principles always have existential implications.[28]

The notion of a nonexistent set is not as odd as it may at first appear, for the existent/nonexistent dichotomy of free set theory is closely akin to the real/virtual distinction of Quine's virtual theory of classes. But while Quine [1963] introduces set abstraction notation by contextual definition and so makes it eliminable — roughly, $t \in \{x \mid A\}$ reduces to $A(t/x)$ — free logics treat set abstracts as genuine singular terms and provide them with an appropriate semantics. Dana Scott [1967] has developed a free positive single-domain model for the virtual theory of classes on which, to a first approximation, the objects of the domain are interpreted as existing sets, and classes of them that are subclasses but not members of the domain are interpreted as nonexisting virtual classes. Scott's semantics validates (Ext) and (Comp); and because on Quine's analysis members of virtual sets must exist, even when the sets do not, it also validates the axiom:

$$s \in t \rightarrow E!s$$

These three axioms form the core of Quine's theory, which may then be augmented with various axioms of existence (pair set, power set, etc.) or nonexistence (e.g., foundation) to yield diverse set theories.

But not all free set theories validate this Quinean principle. Bencivenga ([1976] and [1977]) has developed a theory according to which nonexistent sets can be members of sets, both existing and nonexisting. Like Scott's it begins with the "essential" axioms (Ext) and (Comp). Combining these with free versions of standard set-theoretic axioms Bencivenga ushers us into a strange new set-theoretic universe. Here we discover, for example, that each nonexisting set has an existing singleton — though by (Ext) all these singletons are identical to the null set. This prompts Bencivenga to introduce the additional axiom:

$$t \in \{t\},$$

which gives the null set nonexistent members. Using a version of the foundation axiom (which prohibits self-membership among existing sets) Bencivenga is further able to show that the set of all existing sets does not exist. But it follows from (COMP) and (EXT) that the set of all *non*existing sets does exist and, like singletons of nonexisting sets, is just the null set. Bencivenga suggests that the existing/nonexisting dichotomy of free set theory can do the same job as the set/class distinction of such standard set theories as NGB (von Neumann-Gödel-Bernays) and that the latter may simply be the former misconstrued. But free set theory has so far seen little use or development.

---

[28] One near exception to this is the comprehension principle of Quine's virtual theory of classes — a close relative of free set theory, as will be explained below. But because Quine's logic is not inclusive, even his comprehension principle has existential implications. See Quine [1997].

12 DEFINEDNESS LOGICS AND PROGRAMMING LANGUAGES

For mathematical applications involving *partial functions* (functions whose value may be undefined for some arguments), especially in computability theory, free logic may take the form of a definedness logic.[29] In definedness logics, function symbols, such as the division sign, may represent algorithms, computer programs, terms of a combinatorial logic, partial recursive functions, or the like. Because they deal not with the existence or nonexistence of mathematical *objects*, but with the definedness of operations given various inputs, definedness logics usually employ the postfix definedness predicate '$\downarrow$' instead of the existence predicate '$E!$'. For a given singular term t, $t\downarrow$ is true if and only if the operation specified by t yields some definite value.

What counts as a definite value differs for differing applications. It may, but need not, be required that a computation halts. Thus, for example, on the domain of integers, we may count $(x/y)\downarrow$ as true if and only if $y \neq 0$, though for some values of integers x and y — 1 and 3, for example — the computation of the resulting decimal, though well-defined, never halts.

Though free, definedness logics tend not to be inclusive, since computability theory has little use for the empty domain. Moreover, since it is difficult to think of uncompleted or undefined computations as yielding objects — even nonexistent ones — definedness logics are often of the single-domain variety. These are typically negative.

Yet for application to definedness logics, negative semantics have an important disadvantage: they rule out non-strict functions. (A function is *non-strict* if it may yield a definite value even when its input includes undefined quantities, *strict* otherwise.) Non-strict operators are widely used in programming languages. Lambert notes, for example, that the ternary operator 'if V then x else y' of the programming language ML, which takes a truth-value and two numbers into a number, represents a non-strict function, since it takes the value x when V = true, even if y is undefined.[30] Such operators demand a positive semantics. Note, for example, that from the singular term 'if true then 1 else (1/0)', which denotes the number 1, we may construct the atomic formula:

$$\text{if true then 1 else } (1/0) = 1,$$

which is true even though it contains the empty term 1/0. Yet referentially deficient atomic formulas can be true only in a positive logic.

---

[29] Fefferman [1995] provides an excellent account of definedness logics.

[30] Lambert [2001b], pp. 262–3.

When function symbols are included in the language of a negative free logic, it is necessary (if the logic is to remain negative) to add the axiom schema:

$$f(t_1, \ldots, t_n)\downarrow \rightarrow (t_1 \downarrow \& \ldots \& \ t_n \downarrow),$$

which asserts that f is strict.[31]

Non-strict functions can readily be handled by dual-domain logics. But Lambert suggests that a positive single domain semantics such as the proto-semantics of Antonelli (discussed above) may be preferable, since it avoids introducing mysterious entities whose effect is to make partial or nonstrict functions strict and total.

Yet there is no mystery in using dual-domain semantics to represent error handling. When an undefined term such as '1/0' is encountered in the running of a computer program, the program typically returns an error message. From a semantic viewpoint, these undefined terms may be regarded as denoting error objects, which might, for example, be various error messages. Collectively, these error objects constitute an outer domain, but one that is metaphysically innocuous.[32]

13 MODAL AND TENSE LOGICS

Whereas for many applications the empty terms of free logic have no referent at all, in quantified modal and tense logics they typically do refer, but their referents do not exist in all possible worlds or at all times. The name 'Aristotle', for example, certainly has a referent, but not in possible worlds in which Aristotle was never conceived, nor in the actual world at times before Aristotle's conception or after his death. Thus, taking 'Aristotle' as a, the formula

$$\sim \exists x \ x = a$$

is true in some possible worlds and in the actual world at some times — including the present. But it can be true at all only in a free logic. Hence quantified modal or tense logics are generally free logics.

In modal and tense logics, each world or time has its own domain of objects — the objects that exist in that world or at that time. Since worlds or times at which nothing exists are rarely worth considering, free modal and tense logics are generally noninclusive.

The problem of predication for nonexistents, always an issue for free logics, is especially acute in modal and tense logics. Consider, for example, the statement made at this moment by the present-tense sentence "Aristotle is Greek." Three mutually exclusive conventions seem plausible:

---

[31] This axiom is not a special case of (A−), as Beeson [1985] incorrectly asserts.

[32] For a detailed account of applications of free logic to program specification and verification, see Gumb [2001]. Gumb and Lambert [1991] and [1996] are also helpful.

> This statement is true (Aristotle, though he no longer exists, still satisfies the predicate 'is Greek'),
>
> All such statements are false, because the singular term 'Aristotle' has no currently existing referent, and
>
> All such statements are truth-valueless, for the same reason.

These conventions yield, respectively, positive, negative and nonvalent free logics. Many applications seem to require a positive semantics. Consider, for example, the following present-tense statements, all of which are apparently atomic in form and true:

> Aristotle is famous.
>
> 'Aristotle' refers to Aristotle (where 'Aristotle' is The Philosopher's name in contemporary English).
>
> Aristotle is dead.

In a positive tense logic, however, we must — unless we are willing to abandon the Tarskian analysis of truth in terms of satisfaction — assign to a one-place predicate at a given time an extension that need not be a subset of the domain of that time. (For example, we must include Aristotle in the current extension of the predicate 'is famous', even though Aristotle does not now exist.) Thus we may distinguish for each time an inner domain of objects existing at that time and an outer domain containing objects that may satisfy predicates at that time even though they do not then exist. As with tenses, so with alethic modalities: one-place predicates may have extensions at a world that are not subsets of that world's domain. Hence positive free modal and tense logics tend to employ what is in effect a dual-domain semantics.

A bivalent Kripke model $\mathcal{M}$ for free positive modal or tense logic specifies a set $\mathcal{W}$ of objects representing worlds or times.[33] Each such world or time w has its own domain $\mathcal{D}_w$, the set of things that exist at that world or time, and there is a two-place interpretation function I that assigns denotations to names and extensions to predicates at worlds. The worlds or times are ordered by an accessibility relation $\mathcal{R}$, which expresses relative possibility or temporal succession. Various conditions on $\mathcal{R}$, correspond to various modal logics. Truth is indexed to worlds and/or times, and quantifiers range at a world or time only over that world's or time's domain. Valuation clauses for positive free modal logic (taking the necessity operator '$\Box$' as primitive) generally look something like this:[34]

---

[33] For a more detailed account of Kripke semantics, see Nolt [1997], chs. 12-13.

[34] Clauses for tense logic are essentially the same, except for the tense operators, whose function need not concern us here. If alethic and tense operators are mixed, truth is indexed both to worlds and to times. See Nolt [1997], pp. 365–80, for details.

$\mathcal{V}(Pt_1 \ldots t_n, w) = T$ iff $\langle I(t_1, w), \ldots, I(t_n, w)\rangle \in I(P, w); = F$ otherwise.
$\mathcal{V}(s = t, w) = T$ iff $I(s, w) = I(t, w); = F$ otherwise.
$\mathcal{V}(E!t, w) = T$ iff $I(t, w) \in D_w; = F$ otherwise.
$\mathcal{V}({\sim} A, w) = T$ iff $\mathcal{V}(A, w) = F; = F$ otherwise.
$\mathcal{V}(A \mathbin{\&} B, w) = T$ iff $\mathcal{V}(A, w) = T$ and $\mathcal{V}(B, w) = T; = F$ otherwise.
$\mathcal{V}(\Box A, w) = T$ iff for all u such that $wRu\mathcal{V}(A, u) = T; = F$ otherwise.
$\mathcal{V}(\forall xC, w) = T$ iff for all $d \in \mathcal{D}_w, \mathcal{V}_{(t,d)}(C(t/x), w) = T; = F$ otherwise.

(With regard to the last clause, C is presumed not to contain t and $\mathcal{V}_{(t,d)}$ is the valuation on the model $\mathcal{M}^*$, which is just like $\mathcal{M}$ except that its interpretation function I^* is such that for all $w \in W, I^*(t, w) = d$.[35])

In the least restrictive (and hence most widely applicable) versions of Kripke semantics, domains may vary in size and may, but need not, have common members. Membership of the same object in the domain of more than one world or time models *transworld identity* or *identity over time*. Singular terms that are assigned the same object at every world are called *rigid designators*. Rigid designators (‘Aristotle’ may be an example) may denote objects not contained in the domains of some times (e.g., the present) or worlds (e.g., worlds in which Aristotle’s parents never met). Any logic that combines such rigid designators with quantifiers over those domains must therefore be free.

But *nonrigid designators* (singular terms with world-relative denotation) demand free logics, too, if there are times or worlds at which they fail to denote. Definite descriptions (understood attributively) provide ready examples. ‘The President of the United States’, for instance, is a temporally nonrigid designator that denotes George W. Bush in 2004, Abe Lincoln in 1864, and nothing at times prior to the inauguration of George Washington. Any tense logic that quantifies over the domains of pre-Washington times and treats this description as a genuine singular term must be free. Indeed, short of rejecting objectual quantification over the domains of worlds or times altogether, the only way to make quantified modal or tense logics with singular terms unfree is to impose the exorbitant requirement that each term (rigid or not) have a referent within the domain of each time or world. Freedom is, in general, the better option.

Positive free modal logics with rigid designators may be obtained simply by adding appropriate modal axioms and rules to the standard positive free logic defined by (A1)–(A6) and (A7$^+$). Such logics are, however, too strong for use with nonrigid singular terms. For from $\Box t = t$, which is a theorem of positive free modal logics, and the assumption $E!t$, we obtain $\exists x \Box x = t$. Hence:

$$E!t \rightarrow \exists x \Box x = t$$

---

[35] There are other ways of interpreting quantifiers, but this one, which treats quantificational variables as rigidly designating at w members of $\mathcal{D}_w$, is perhaps the simplest and most generally useful.

is a theorem of positive free modal logic. But this formula is valid only where t is rigid. It follows that the quantifier rules must be weakened, and perhaps other adjustments made, in positive modal logics that allow nonrigid designators.[36]

For negative or nonvalent free modal or tense logics, predicates are interpreted at a world or time only over the domain of that world or time, but singular terms may still denote objects that at a given world or time do not exist. The resulting semantics is therefore not single-domain (since singular terms may have referents outside a world's or time's quantificational domain), but neither is it typically dual-domain, for it does not permit objects outside that domain of quantification to satisfy predicates at that world or time.

Such negative and nonvalent logics thus separate two roles that in a non-modal dual-domain semantics are both played by the outer domain: being a reservoir of referents for empty terms and bounding the extensions of predicates. These two roles may be separated in positive logics as well. For there are at least three kinds of structures that might constitute an outer domain on a Kripkean modal semantics, and each of these might play either role:

(i) The outer domain might, like the inner, be world-relative, consisting for a given world w of the union of the domains of all the worlds accessible from w; hence it would include all objects possible relative to w.

(ii) The outer domain might be the same for all worlds, consisting of the union of domains of all the worlds. Such a domain would include all objects that are possible relative to any world

(iii) The outer domain might be a proper superset of the union of domains of all the worlds — and hence include even impossible objects.

It is pointless to make the domain of predicate extensions wider than both the domain of referents and the domain of quantification (since in that case the extra "width" of the predicate extensions has no effect on truth), but any other combination is potentially useful. Thus we might, for example, draw predicate extensions at w only from within the union of the domains of worlds accessible from w but allow singular terms to denote "impossible" objects that exist in the domain of no world. If t were to denote one of these impossibilia, then, assuming bivalence, any atomic formula containing t would have to be false at every world. Such a semantics would validate the principle:

$$Ft_1\ldots t_n \rightarrow (\Diamond E!t_1 \ \& \ \ldots \ \& \ \Diamond E!t_n), \text{ where } F \text{ is primitive,}$$

a modal analog of (A$-$).

Restricting the referents of rigidly designating singular terms to the sort of outer domain defined in convention (i) would raise the issue of *which* outer domain we were talking about — since our language applies to all worlds, and under this convention each world has its own outer domain. This question could be answered in a Kripkean semantics by designating a particular world as actual, for then the actual world's outer domain can be regarded as *the* outer domain for purposes of referring. (With a designated actual world we also have a non-world-relative concept of truth — namely, truth in the actual world, and validity is defined as truth *at the actual world* in all models.) Such a semantics requires singular terms to denote objects that are actually possible and hence validates the axiom $\Diamond \mathrm{E!}t$. But unless all worlds are assumed accessible from the actual world, it has the disadvantage of permitting quantification in worlds inaccessible from the actual world over things that cannot be named. Still, convention (i) has applications, as we will see in the next section.

---

[36] See Garson [1991], pp. 120-1 for an account of some of the options.

If we define the domain of referents using convention (ii), then there is only one such outer domain (whether we specify an actual world or not), and it is the same for all worlds. This domain contains any object that exists in any possible world. If we require the accessibility relation to be universal (so that the logic is S5) — then, once again, the schema $\Diamond \mathrm{E!}t$ is valid for every term t.

Finally, if we define the domain of referents by convention (iii), $\Diamond \mathrm{E!}t$ will be invalid, regardless of how we restrict accessibility. Thus by mixing and matching different kinds of outer domain we can produce various logics. More variety can be added if we consider truth value gaps and supervaluations, quantification over individual concepts, and the like — but enough is enough.[37]

14 INTUITIONISTIC LOGICS

In intuitionistic mathematics an object is deemed to exist only if it is known to be constructible by some definite process. Yet the standard intuitionistic logic, the Heyting predicate calculus (HPC), validates $\exists x\ x = t$ for every singular term t. The proper applicability of HPC is therefore restricted to statements whose singular terms all denote objects that are known to be constructible.

Yet, remarkably, the Kripke semantics originally developed for HPC, on which HPC has repeatedly been "proved" sound and complete, actually renders HPC *un*sound, for it provides counterexamples to the classical quantifier rules and validates instead a free intuitionistic predicate logic.[38] The errors in several of the published soundness proofs for HPC arise from a failure of the proffered versions of Kripke semantics to guarantee that singular terms always denote objects that are known to be constructible. To see why, we must take a glance at Kripke semantics for intuitionistic logic.

---

[37] For treatment of these topics, see Garson [1991]. Other useful sources include: Cocchiarella [1991], Schweitzer [2001] and Simons [2001]. Barba [1989] provides a supervaluational semantics for modal free logic.

[38] See Posy [1982], p. 52 and endnotes 4 and 6, and Leblanc and Gumb [1983], footnote 31, p. 194. Posy and Leblanc and Gumb observe that the soundness proofs of both Kripke [1963] and Thomason [1968] err in this way, though Posy mistakenly gives 1965 as the date of Thomason's article. The same error occurs in Troelstra and van Dalen [1988].

Kripke semantics is bivalent, the values T and F expressing (for mathematical applications, at least) the ideas "proved" and "unproved," respectively. As in Kripke's semantics for modal logic, each Kripke model contains a set $\mathcal{W}$ of "worlds," which in intuitionistic models are best understood as states of mathematical knowledge. Each world has its associated domain $\mathcal{D}_w$ (intuitively, the set of all objects known to be constructible at w). A two-place interpretation function I assigns denotations to singular terms and extensions to predicates at worlds. (Predicate extensions may be drawn just from the domain of the world itself or, for positive free logics, from some wider "outer" domain; see the discussion of outer domains in the previous section.) I then induces, as usual, a valuation assignment $\mathcal{V}$ from formulas and worlds to truth-values. The worlds are ordered by a reflexive, transitive accessibility relation $\mathcal{R}$.[39] Intuitively, wRu signifies that u is epistemically possible relative to w — in other words, that whatever is proved in w, and possibly more, is proved in u. We therefore stipulate that if $w\mathcal{R}u$, then:

> $\mathcal{D}_w \subseteq \mathcal{D}_u$,
> for all sentence letters P, if $I(P,w) = T$, then $I(P,u) = T$, and
> for all n-place predicates $F(n \geq 1), I(F,w) \subseteq I(F,u)$.

Objects contained in the domains of worlds accessible from w are objects whose construction is epistemically possible at w — objects, in other words, that might be constructible for all that is known in w.

It is not necessary for our purposes to discuss the semantics of the propositional operators, but we must consider the semantics of the quantifiers. While the existential quantifier behaves as usual, the universal works like '$\Box\forall$'. More precisely, let $\mathcal{M}$ be a Kripke model with interpretation function I, and let t be any individual constant t not occurring in A. Then:

> $\mathcal{V}(\exists xC, w) = T$ iff for some $d \in \mathcal{D}_w, \mathcal{V}_{(t,d)}(A(t/x), w) = T; = F$ otherwise.
> $\mathcal{V}(\forall xC, w) = T$ iff for all u such that $w\mathcal{R}u$ and all $d \in \mathcal{D}_u, \mathcal{V}_{(t,d)}(A(t/x), u) = T; = F$ otherwise.

(Here, as in the previous section, C is presumed not to contain t and $\mathcal{V}_{(t,d)}$ is the valuation on the model $\mathcal{M}^*$, which is just like $\mathcal{M}$ except that its interpretation function I^* is such that for all $w \in W, I^*(t,w) = d$.)

Whether the logic is free or not depends on what singular terms may denote. If we require models to specify an actual world, treat truth as truth at that world, and restrict denotation to objects within that world's domain (intuitively, the objects known to be constructible), then the resulting logic is HPC, which is not free. But most versions of Kripke semantics (including those designed for HPC on which the erroneous soundness proofs are based) do not thus restrict denotation. They allow singular terms to denote either all objects in the domains of worlds accessible from the actual world (which, of course, is possible only in that models specify an actual world), all objects in the union of domains of all the worlds, or all objects in some still larger set that includes this union. These three options correspond, respectively, to delimiting the domain of referents by conventions (i), (ii) or (iii) of the previous section. In all three cases, the resulting logic is free.

---

[39] Sometimes an antisymmetry requirement is added.

This may seem implausible for convention (i). Given the semantic difference between universal and existential quantifiers, it might seem that if singular terms denoted only objects in the domains of worlds accessible from the actual world (convention (i)), these objects would all be within the domain of the universal quantifier but perhaps not within the domain of the existential, so that the resulting logic would validate specification but not existential generalization. But here the presumed extent of the universal quantifier's domain is mistaken. Despite the universal quantifier's exotic semantics, its domain at w is just $\mathcal{D}_w$. If term t denotes an object that does not exist at w (i.e., is not in $\mathcal{D}_w$) but does exist at some world u accessible from w and $\forall xA$ is true at w, then although $A(t/x)$ must be true at u, it needn't be true at w. The situation is, then, as I noted in the previous paragraph: even a convention-(i) semantics yields a free logic in which both classical specification and classical existential generalization are *in*valid.

Carl Posy [1982] has argued that such free intuitionistic logics are more "natural" than HPC. Whether or not they more natural, they certainly are more widely applicable; for since they do not validate $\exists x\ x = t$, their singular terms need not denote only objects that are known to be constructible.

Free intuitionistic logics also exhibit the usual variations arising from the differences between positive, negative and nonvalent semantics and some less usual variations arising from differing conventions regarding predication. Posy [1982] provides a thorough account of these.

15 FAILURE OF SUBSTITUTIVITY OF THE BICONDITIONAL[40]

Despite its utility for the applications just considered — and others not considered — freedom has its price. One disadvantage of all reasonably strong free logics — positive, negative or nonvalent — is that they violate the principle of substitutivity of the biconditional. They do not, in other words, permit coextensive open formulas always to be substituted for one another *salva veritate.* (Two open formulas in n free variables are *coextensive* if every n-tuple of members of the quantificational domain satisfies one if and only if it satisfies the other.) This anomaly results from the fact empty singular terms may yield different truth-values when substituted for the free variables of coextensive open formulas.[41] Consider the substitution of an empty term t for x in the three coextensive open formulas $x = x, (x = x \ \& \ E!x)$, and $(E!x \rightarrow x = x)$. A positive free logic makes $t = t$ true and $(t = t \ \& \ E!t)$ false, a negative free logic makes $t = t$ false and $(E!t \rightarrow t = t)$ true, and any reasonably strong nonvalent free logic — NINO logics do not count as reasonably strong — makes $t = t$ truth-valueless and $(E!t \rightarrow t = t)$ true.

---

[40] This section is based on Lambert [2003e] and [2003g]. Lambert suggests a way of avoiding failure of substitutivity by analysis of the notion of predication.

[41] In an E-logic, of course, the variables need not be assigned members of the quantificational domain, but that is irrelevant here, since we are concerned with coextension.

The loss is greatest in single-domain logics. Dual-domain logics provide some compensation, for although their coextensive expressions are likewise not intersubstitutible *salva veritate*, their *co-comprehensive* expressions (i.e., those with the same satisfaction conditions over the union of the inner and outer domains) are. Co-comprehensiveness, however, is not expressible in an object language whose formulas are closed,[42] unless we introduce quantifiers over the outer domain — a move that free logicians have generally opposed. But we have already met such quantifiers in connection with Grandy's theory of definite descriptions, and the next section bolsters the case for their use.

16 AN EXPRESSIVE LIMITATION

Free logic, unlike classical logic, enables us to express contingent existence assertions about particular individuals. This raises the hope that it might profoundly illuminate philosophical arguments for individual existence. Not so. Consider, for example, this simplistic and obviously valid (but not obviously sound) version of the ontological argument:

God has all perfections.
Whatever has all perfections exists.
∴ God exists.

Classical logic accounts this argument valid — but irrelevantly, since its validity stems from the logical truth of its conclusion. But in free logic, it is not even valid — at least if we construe the second premise as $\forall x(Px \rightarrow E!x)$. For this formula is in free logic (as in classical logic with E!x expressed as $\exists y\ y = x$) contentless — a mere logical truth. To infer from the second premise by specification the instance:

If God has all perfections, God exists,

which is needed to deduce the conclusion, we must in free logic add the question-begging premise 'God exists'.

This is not an isolated problem. *Every* statement of sufficient conditions for existence (e.g., 'whatever thinks exists', 'anything that can be photographed exists', 'whatever interacts causally with existing things exists' — even the seemingly absurd 'every conceivable object exists') is vacuously *true*, so long as the quantifier ranges only over existing things. (It is, in fact, axiomatic by (A5) that *everything* exists.) As a premise, any statement of sufficient conditions for existence is therefore otiose. To express existence conditions that may be false and hence contribute substantively to an argument — we need quantifiers ranging over nonexisting as well as existing things.

---

[42] It is, however, expressible in an E-logic.

This is easily achieved in dual-domain frameworks: just let the quantifiers range over the union of the inner and outer domains. But because that divorces quantification from existence and offends tastes for desert landscapes, early attempts to introduce quantifiers over nonexistents were often summarily dismissed.[43] Growing appreciation of their usefulness has, however, made the recent literature more sympathetic.[44] It is now almost respectable to maintain that conflating "for at least one" with "there exists" was a mistake. Alex Orenstein [1990], for example, argues that the only way to account for the logical connection between generalized conjunctions and what are commonly called "existential" quantifiers — he prefers to call them "particular," which neatly captures the duality with "universal" — is to recognize that the latter do not express existence. Richard Routley goes even further, insisting that free logic is just one of several steps in the elimination of existential presuppositions — a project that will culminate only when quantifiers are at last wholly free to range over domains containing nonexistents.[45]

In the section on definite descriptions I introduced the generalized product symbol 'Π' for the universal quantifier over the outer domain of referents. Here let's add the generalized sum symbol 'Σ' for its "existential" dual. These quantifiers provide enough expressive power to handle typical arguments for the existence of individuals. In particular, the simplistic ontological argument at the beginning of this section becomes:

$$Pg, \Pi x(Px \rightarrow E!x) \vdash E!g,$$

which is relevantly valid.

Since singular terms no longer denote anything outside the domain of quantification, the resulting logic is not free but classical. But, taking '$E!$' as primitive, we can define a separate set of free quantifiers, restricted to the domain of existing things, as follows:

$$\begin{aligned} \forall x A &=_{df} \Pi x(E!x \rightarrow A) \\ \exists x A &=_{df} \Sigma x(E!x \ \& \ A) \end{aligned}$$

These quantifiers comply with (A1)–(A5), preserving in a wider context the free logic with which we began.

But now comes a new epicycle and a new question. For suppose that the outer domain is the set of all *possible* objects. Then attempts to express sufficient conditions for *possibility* will, like our previous attempts to express sufficient conditions for *existence*, yield only vacuity. So we might decide to quantify classically over a still wider outer domain that includes *im*possible objects, add an analog of '$E!$' that expresses possibility rather than existence, and introduce by definition restricted quantifiers over the old domain of *possible* objects. The logic of these defined quantifiers over possibilia is formally identical to a free logic. Does that make it one?

---

[43] Lejewski [1954] or Routley [1966], for example.

[44] Quantifiers that range over possible objects that need not actually exist have played an increasingly prominent role in modal and intensional logics. Even broader nonactual domains are subject to quantification in the Meinongian literature mentioned in note 8.

[45] Routley [1980], pp. 77-81.

17 WHAT IS A FREE LOGIC?

Not by the standard definitions, which identify the domain of quantification with the set of *existing* things. Bencivenga, for example, stipulates that

> [a] free logic is a formal system of quantification theory, with or without identity, which allows for some singular terms in some circumstances to be thought of as denoting no existing object, and in which quantifiers are invariably thought of as having existential import.[46]

Morscher and Simons give this definition:

> A logical system L is a *free logic* iff
>
> 1. L is free of existential presuppositions with respect to the singular terms of L,
> 2. L is free of existential presuppositions with respect to the general terms of L, and
> 3. the quantifiers of L have existential import.[47]

And Lambert offers this compact equivalent: "The expression 'free logic' is an abbreviation for the phrase 'free of existence assumptions with respect to its terms, general and *singular*'."[48]

Note, incidentally, that these definitions appear *not* to be equivalent. Morscher and Simons' and Lambert's, unlike Bencivenga's, require *general* terms not to have existential import. This would seem to make all free logics inclusive, since, as I noted above, universal general terms have existential import in nonempty domains. But in fact all that these authors mean in denying existential import to general terms (or predicates) is that general terms may be empty — a point that all modern logicians take for granted. The definitions, as intended, are therefore all equivalent.

Furthermore, to return to my main point, these authors all agree that in free logics, at least, the domain of quantification is the set of existing things. But the discussion of the previous section suggests the following purely semantic characterization of positive free logics: they are fragments of classical logics obtained by restricting the quantifiers to a proper subset of the original domain. (Negative or nonvalent free logics cannot be so characterized, since they make self-identity statements untrue — a classical impossibility.) More precisely, consider a positive free logic axiomatized by (A1)–(A6) and ($A7^+$), whose primitive operators are '$\sim$', '$\rightarrow$', '$\forall$' and '$E!$' and a classical logic using the same predicates, singular terms and primitive operators, except that it has a new monadic predicate 'E' in place of '$E!$'. Now define a translation *Tr* from the free logic to the classical logic as follows:

---

[46] Bencivenga [1985], p. 375.

[47] Morscher and Simons [2001], p. 2.

[48] Lambert [2003f], p. 123.

$$\begin{aligned}
&Tr(E!t) = Et\\
&Tr(Ft_1 \ldots t_n) = Ft_1 \ldots t_n, \text{ where } F \text{ is any other primitive predicate,}\\
&\qquad \text{including `} =' \\
&Tr({\sim}A) = \ {\sim}Tr(A)\\
&Tr(A \to B) = Tr(A) \to Tr(B)\\
&Tr(\forall xA) = \forall x(Ex \to Tr(A)).
\end{aligned}$$

Then for any formula A, A is a theorem of the free logic if and only if $Tr(A)$ is a theorem of the classical logic.[49] This translation suggests a purely semantic conception of positive free logic: it is just that part of classical logic whose quantifiers are restricted to a special subset of the domain — intuitively, the set of "existing" things. Freedom, on this conception, is restriction.

The translation, incidentally, illuminates the connection between free and inclusive logics, for emptiness of the domain of the free logic is represented in the classical logic as the emptiness of the predicate 'E'. The admissibility of the empty domain is thus reduced to the fact that in classical predicate logic general terms lack existential import.

Should we adopt this purely semantic characterization of positive free logic? Clearly not, insofar as it utilizes classical quantification over an outer domain. Those who reject either dual-domain semantics or quantification over nonexistent entities may justly dissent. Yet a purely semantic characterization of free logic (whether single-domain or dual-domain; positive, negative or nonvalent) is possible — and, indeed, was given at the outset of this essay: free logics are logics whose singular terms need not denote members of the quantificational domain. (If there are two or more quantificational domains, as in dual-domain logics with quantifiers over the outer domain, we must specify which domain and which quantifiers we are talking about; a logic may be free with respect to some but not all of its quantifiers.) This definition has the advantage of replacing a metaphysical criterion (the terms of a free logic need not denote *existents*) by a clear semantic one (they need not denote members of the quantificational domain), and on it the logic of possibilia described in the previous section is free.

18 CONCLUSION

Though unsullied by existential commitment, free logic does not reveal a tidy and compelling realm of logical truth. In fact, the whole business is disappointingly messy. Each new application requires new semantic decisions, and over each decision pragmatic considerations hold sway: Should we allow empty domains? Should we adopt a single-domain or dual-domain approach? If the latter, how wide should we make the outer domain of referents, should we distinguish it from the outer domain of predicates, and what identity conditions should we ascribe to its nonexistent objects? Should we quantify over these objects? If free variables are allowed, may they take values outside the quantificational domain or not? Should the semantics be positive, negative or nonvalent? If positive, by what conventions shall we assign truth-values to referentially deficient formulas? If nonvalent, which valuation rules should we choose, and should we base our logic on strong or weak validity? If there are truth-value gaps, should we augment the semantics with supervaluations? And so on ... . Freedom does not extricate us from the general flux of ideas; it only creates new eddies within the flux.

---

[49] This was first shown by Meyer and Lambert [1968]. Bencivenga [1985], pp. 412-14, offers a helpful summary. See also Trew [1970].

Nor have free logics yielded profound philosophical insights into reasoning about the existence of particular things. They are, however, handier than classical logic for many applications and practically indispensable for some — especially modal and tense logics, where they have made reasoning about possible but nonactual or past and future objects a good bit more intelligible. As regards the realm of the nonexistent (where some would say there is nothing to learn), they have stimulated the investigation of semantic conventions that can generalize, unify and simplify theory. They sometimes enable us to separate questions of "essence" from those of existence, as in free set theory. And finally, they have significantly advanced our understanding of the semantics of definite descriptions and facilitated a host of more modest advances in other areas.

In logic, as elsewhere, freedom, though messy, is often desirable.[50]

BIBLIOGRAPHY

[Antonelli, 2000] G. A. Antonelli. Proto-semantics for positive free logic, *Journal of Philosophical Logic* 29,3, pp. 277-94, 2000.

[Beeson, 1985] M. Beeson. *Foundations of Constructive Mathematics*, Berlin: Springer-Verlag, 1985.

[Barba, 1989] J. L. Barba. A modal version of free logic, *Topoi* 8, pp. 131-5, 1989.

[Bencivenga, 1976] E. Bencivenga. Set theory and free logic, *Journal of Philosophical Logic* 5, pp. 1-15, 1976.

[Bencivenga, 1977] E. Bencivenga. Are arithmetical truths analytic: new results in free set theory, *Journal of Philosophical Logic* 6, pp. 319-30, 1977

[Bencivenga, 1981] E. Bencivenga. Free semantics, in *Boston Studies in the Philosophy of Science* 47, pp. 38-41, 1981. Revised version reprinted in [Lambert, 1991, 98-110].

[Bencivenga, 1985] E. Bencivenga. Free Logics, in D. Gabbay and F. Guenthner, eds., *Handbook of Philosophical Logic*, vol. III: *Alternatives to Classical Logic*, Dordrecht: D. Reidel, 1985.

[Burge, 1974] T. Burge. Truth and singular terms, *Noûs* 8, pp. 309-25, 1974; reprinted in [Lambert 1991, 189-204].

[Cocchiarella, 1991] N. B. Cocchiarella. Quantification, time and necessity, in [Lambert, 1991, 242-256].

[Feferman, 1995] S. Feferman. Definedness, *Erkenntnis* 43,3, pp. 295-320, 1991.

[Fine, 1983] K. Fine. The permutation principle in quantificational logic, *Journal of Philosophical Logic* 12, pp. 33-7. 1983.

---

[50] I wish to thank the Logic Reading Group at the University of Tennessee, Knoxville — Richard Gale, Jim Okapal, Christopher Pynes, and Jake Sumner — for helpful suggestions and corrections to this paper and Sarah Kenehan for extensive help with the research.

[Garson, 1991] J. W. Garson. 'Applications of free logic to quantified intensional logic, in [Lambert, 1991 111-42].

[Grandy, 1972] R. E. Grandy. A definition of truth for theories with intensional definite description operators, *Journal of Philosophical Logic* 1, pp. 137-55, 1972. Reprinted in [Lambert, 1991, 171-88].

[Gumb, 2001] R. D. Gumb. Free logic in program specification and verification, in [Morscher and Heike, 2001, 157-93].

[Gumb and Lambert, 1991] R. D. Gumb and K. Lambert. Definitions in nonstrict positive free logic (corrected), Modern Logic, 7, 25-55 and 435-440 (errata), 1991.

[Gumb and Lambert, 1996] R. D. Gumb and K. Lambert. A free logical foundation for nonstrict functions, Proceedings of the CADE-13 Workshop on the Mechanization of Partial Functions, pp. 39-46, 1996.

[Hintikka, 1959] J. Hintikka. Towards a theory of definite descriptions. *Analysis* 19, pp. 79-85, 1959.

[Jacquette, 1996] D. Jacquette. *Meinogian Logic: The Semantics of Existence and Nonexistence*, Berlin: Walter de Gruyter, 1996.

[Kripke, 1963] S. Kripke. Semantical analysis of intuitionistic logic I, in J. Crossley and M. Dummett, eds., *Formal Systems and Recursive Functions*, Amsterdam, North Holland, pp. 92-129, 1963.

[Lambert, 1983] K. Lambert. *Meinong and the Principle of Independence*, Cambridge: Cambridge University Press, 1983.

[Lambert, 1991] K. Lambert, ed. *Philosophical Applications of Free Logic*, New York: Oxford University Press, 1991.

[Lambert, 2001a] K. Lambert. Free logic and definite descriptions, in [Morscher and Heike, 2001, 37-47].

[Lambert, 2001b] K. Lambert. From predication to programming, *Minds and Machines* 11,2, pp. 257-65, 2001.

[Lambert, 2003a] K. Lambert. *Free Logic: Selected Essays*, Cambridge: Cambridge University Press, 2003.

[Lambert, 2003b] K. Lambert. Existential import, E! and 'The' in [Lambert, 2003a, 16-32].

[Lambert, 2003c] K. Lambert. Foundations of the hierarchy of positive free definite description theories, in [Lambert, 2003a, 69-91].

[Lambert, 2003d] K. Lambert. The Hilbert–Bernays theory of definite descriptions, in [Lambert, 2003a, 44-68].

[Lambert, 2003e] K. Lambert. Nonextensionality, in [Lambert, 2003a, 107-21].

[Lambert, 2003f] K. Lambert. The philosophical foundations of free logic. in [Lambert, 2003a, 122-75].

[Lambert, 2003g] K. Lambert. Predication and extensionality, in [Lambert, 2003a, 92-106].

[Lambert, 2003h] K. Lambert. Russell's version of the theory of definite descriptions, in [Lambert, 2003a, 1-15].

[Leblanc, 1971] H. Leblanc. Truth value semantics for a logic of existence, *Notre Dame Journal of Formal Logic* 12, pp. 153-68, 1971.

[Leblanc and Gumb, 1983] H. Leblanc and R. D. Gumb. Soundness and completeness proofs for three brands of intuitionistic logic, in *Essays in Epistemology and Semantics*, Leblanc, Gumb and Stern, eds., New York, Haven, 1983

[Leblanc and Thomason, 1968] H. Leblanc and R. H. Thomason. Completeness theorems for some presupposition-free logics, *Fundamenta Mathematicae* 62, pp. 125-64. Reprinted in Leblanc's *Existence, Truth and Provability*, Albany: State University of New York Press, 1982, pp. 22-57, 1968

[Lehman, 1994] S. Lehman. Strict Fregean free logic, *Journal of Philosophical Logic* 23,3, pp. 307-336, 1994.

[Lehman, 2001] S. Lehman. No input, no output logic, in [Morscher and Hieke, 2001, 147-54].

[Lejewski, 1954] C. Lejewski. Logic and existence, *British Journal for the Philosophy of Science* 5, 18, pp. 104-19, 1954. Reprinted in Dale Jacquette, ed., *Philosophy of Logic: An Anthology*, Oxford, Blackwell, 2002, pp. 147-55.

[Meyer *et al.*, 1982] R. K. Meyer, E. Bencivenga and K. Lambert. The ineliminability of E in free quantification theory without identity, *Journal of Philosophical Logic* 11, pp. 229-31, 1982.

[Meyer and Lambert, 1968] R. K. Meyer and K. Lambert. Universally free logic and standard quantification theory, *Journal of Symbolic Logic* 33, pp. 8-26, 1968.

[Morscher and Hieke, 2001] E. Morscher and A. Hieke, eds. *New Essays in Free Logic: In Honour of Karel Lambert* (Applied Logic Series, vol., 23), Dordrecht: Kluwer, 2001.

[Morscher and Simons, 2001] E. Morscher and P. Simons. Free logic: a fifty-year past and an open future, in [Morscher and Hieke, 2001].

[Nolt, 1997] J. Nolt. *Logics*, Belmont, CA: Wadsworth, 1997.

[Orenstein, 1990] A. Orenstein. Is existence what existential quantification expresses? in Robert B. Barrett and Roger F. Gibson, eds., *Perspectives on Quine*, Cambridge: Blackwell, pp. 245-70, 1990.

[Paśniczek, 1998] J. Paśniczek. *The logic of intensional objects*: *A Meinongian Version of Classical Logic*, Dordrecht: Kluwer, 1998.

[Paśniczek, 2001] J. Paśniczek. Can Meinongian logic be free? in [Morscher and Hieke, 2001, 227-36].

[Posy, 1982] C. J. Posy. A free IPC is a natural logic: strong completeness for some intuitionistic free logics, *Topoi* 1, pp. 30-43, 1982. Reprinted in [Lambert, 1991, 49-81].

[Quine, 1954] W. V. O. Quine. Quantification and the empty domain, *Journal of Symbolic Logic* 19, pp. 177-9, 1954.

[Quine, 1963] W. V. O. Quine. *Set Theory and its Logic*, Cambridge: Harvard, 1963.

[Quine, 1997] W. V. O. Quine. Free logic, description and virtual classes, *Dialogue* 36,1, pp. 101-8, 1997.

[Routley, 1966] R. Routley. Some things do not exist, *Notre Dame Journal of Formal Logic* 7, pp. 251-76, 1966.

[Routley, 1980] R. Routley. *Exploring Meinong's Jungle and Beyond*, Canberra: Australian National Unversity, 1980.

[Russell, 1919] B. Russell. *Introduction to Mathematical Philosophy*, New York: Simon & Schuster, 1919.

[Schweitzer, 2001] P. Schweitzer. Free logic and quantification in syntactic modal contexts, in [Morscher and Hieke, 2001].

[Scott, 1967] D. Scott. Existence and description in formal logic, in Ralph Schoenman, ed., *Bertrand Russell, Philosopher of the Century*, London, Allen and Unwin, pp. 181-200, 1967. Reprinted in [Lambert, 1991, 28-48].

[Scott, 1979] D. Scott. Identity and existence in intuitionistic logic,' in Fourman, M. P., Mulvey, C. J. and Scott, D. S., eds., *Applications of Sheaves*, Lecture Notes in Mathematics 753, Berlin: Springer-Verlag, pp. 660-696, 1979.

[Simons, 2001] P. Simons. Calculi of names: free and modal, in [Morscher and Hieke, 2001].

[Thomason, 1968] R. Thomason. On the strong semantical completeness of the intuitionistic predicate calculus, *Journal of Symbolic Logic* 33, pp. 1-7, 1968.

[Trew, 1970] A. Trew. Nonstandard theories of quantification and identity, *Journal of Symbolic Logic* 35, pp. 267-94, 1970.

[Troelstra and van Dalen, 1988] A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics*, vol. 1 Amsterdam: North Holland, 1988.

[van Fraassen, 1966] B. C. van Fraassen. Singular terms, truth value gaps and free logic, *Journal of Philosophy* 63, pp. 481-95, 1966. Reprinted in [Lambert, 1991, 82-97].

[Woodruff, 1984] P. W. Woodruff. On supervaluations in free logic, *Journal of Symbolic Logic* 49, pp. 943-50, 1984.

FICTIONS AND THEIR LOGIC

John Woods*

1 LOGIC

The logic of fiction came into its own in the 1970s with the publication of a number of seminal works.[1] Fiction has been a stand-alone research programme for logicians only since that time. The logician's interest in fiction had a number of early motivations. However, in a quite general way, the logic of fiction programme arose out of failure. Fictional discourse simply wouldn't fit existing systems. As Richard Routley once put it, "literary phenomena rather convincingly show the inadequacy of most going formal semantics." [Routley, 1979, 3] In particular, the sentences of fiction have not been well-handled in the more traditional theories of nondenoting terms, such as description theory, free logic or supervaluational semantics. On the other hand, fictional discourse appears to be tailor-made for existence-neutral theories of quantification. A third motivation is fictionalism in the philosophy of mathematics. If mathematical objects are to be likened to fictional objects, then it would seem to be incumbent on those who so propose that they give a disciplined account of what it is to be an object of fiction. A fourth consideration is that since it would appear that some fictional truths are inconsistent, this may be some reason

---

[*] [Woods and Alward, 2002] reflects my thinking about the fictional some thirty years after my original work in the period 1969 to 1974. In writing it, Alward and I brought different inclinations to bear. For Alward, these would be first words in a continuing project on fiction. For me, they would be last words, a kind of valedictory. Alward has turned out to be more prescient than I. He has extended his thinking considerably, and a book will appear in a year or so [Alward, forthcoming]. But, in what can only be considered a surprising turn, I too have continued to think about fiction, and I too have drifted from where I was in 2002. I shall mention just three of the newer developments. The 2002 paper gives no detailed attention to what it would take for fiction to have a *logic*, as opposed to a philosophical analysis of the kind a philosopher could furnish. In the present work, efforts have been made to repair this omission. In the 2002 chapter, stipulationism is flirted with. Here it is embraced fully. In 2002, I leaned toward the theoretical primacy of mathematical stipulationism over fictional stipulationism (with Alward demurring, but not in print.) I have now come to my senses. Mathematical stipulationism is a harder nut to crack than literary stipulationism. Hats off to Alward.
For expository ease, I have in a few places drawn modestly upon [Woods and Alward, 2002], with Alward's and Kluwer's kind permission. Aside from these expository matters, the present chapter is a considerable departure from earlier work. Errors are mine alone. Alward bears no responsibility for them. He will make his own in [Alward, forthcoming].

[1] These include in chronological order [Woods, 1969; Gale, 1971; Routley, 1973; Coleman, 1973; Crittenden, 1973; Woods, 1974; Devine, 1974; Martin, Schotch, 1974; Blocker, 1974; Howell, 1974; Pavel, 1975; Gabriel, 1975; Parsons, 1975; Woltersdorff, 1976; Howell, 1976; Urmson, 1976; Lewis, 1978; Walton, 1978a, 1978b; Parsons, 1978; Woods and Pavel, 1979]. In addition to works expressly designed to accommodate the analysis of fictional discourse, various other contributions of the period bear on the issue somewhat tangentially, but with interesting and influential things to say about fiction [Kripke, 1972; Kaplan, 1973; Plantinga, 1974].

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

to construe logic dialetheically. Other motivations concern philosophy of language and epistemology rather more centrally, but some of these also have a bearing on modal logic, especially theories that postulate fictive operators. A case in point is the apparent inability of widely received theories of reference and of knowledge to account for our capacity to refer to fictional objects and to know something of them. Such theories are causal in their orientation, and they to give rise to problems not unlike Benacerraf's Dilemma in the philosophy of mathematics [Benacerraf, 1973], which offers us — so it is said — a difficult choice between "easy epistemology for unbelievable objects or implausible epistemology for constructed objects". [Thomas, 2002, 19].[2] We shall return to these pre-existing logics in §3.5 below

The first issue for the aspiring logician of fiction is what a logic of fiction would be. What would it be like for fiction to *have* a logic, and in what sense of the word 'logic'? Bearing on this is a distinction adumbrated — albeit not under this name — in the writings of Frege. It is the distinction between a *mathematical* and a *philosophical* logic. What we find in the *Begriffsschrift* [1967] is mathematical logic. What we find in "Sense and Reference" [1952] is philosophical logic. A philosophical logic investigates properties such as reference, truth and inference in natural languages. Accordingly, there is a substantial overlap between philosophical logic and the philosophy of language.

Reference figures significantly in Frege's mathematical and philosophical logics alike. In his mathematical logic, it is a simplistic and rather naïve concept. The reference of names is represented by a primitive notion of object-designation. The reference of individual variables is relativized to countable sequences of arbitrary objects; a variable v_n bears this primitive relation of denotation to the object occupying the nth place in that sequence, where 'n' is the variable's index. Predicate symbols are taken to denote sequences of abstract objects. Plural reference is seen as the quantificational binding of variables.

In reading Frege's papers on the philosophy of language, one is struck at once by the sophistication of its treatment of reference. While the theory retains a denotational cast (in which referring, crudely speaking, is like labelling), considerable attention is paid to semantic constraints, that is, to the role played by (something like) a referring term's meaning. Frege's analysis of these semantic constraints focuses on the notion of *Sinn*. But there is no place for *Sinn* in Frege's mathematical logic.

PROPOSITION 1 (Underdetermination). *Frege's mathematical logic of reference underdetermines his philosophical logic of reference.*

Underdetermination as such is nothing to get over-excited about. Even our best formal models always leave some things out of account. What matters is whether reference is well-enough handled to enable the logic to elucidate its own target properties, such as entailment and logical truth. Another way of saying this is that reference is not a primary objective for standard approaches to mathematical logic. It is more an instrumental necessity. One has to say *something* about reference in order to be able to say what one *wants* to say about entailment. If we wished to have a full-blown conceptual analysis of reference, it is clear that we would not find it in Frege's mathematical logic. We would need to explore the relevant options in the philosophy of logic. If we came to be satisfied that the right account could be found there, a further possibility would present itself. We might return to some preferred mathematical logic in an effort to determine whether, under suitable adaptation, the philosophical account of reference could be accommodated by the logic's formalisms, or some natural enrichment of them.

---

[2] It is widely held that the best theory of knowledge to date is one in which knowledge arises from the knower's causal contact with the object of his knowledge (albeit, on occasion, a quite complex sort of causal contact). It is also widely held that the best theory of truth to date is one in which the truth of a sentence, e.g., "Harry is a man" consists of the object denoted by "Harry" being an element in the class of objects denoted by "is a man". Similarly, "2 is a prime" would be true if and only if the number two were a member of the set of prime numbers. The gist of Benacerraf's Dilemma is this: If the causal theory of knowledge is right, it is difficult to see how mathematical knowledge is possible. For we seem not to have causal contact with such objects.

Mathematical logics put a premium on mature and coherent formalization. The devil is in the details, of course, but in a general way it can be said that formalizations are well-specified uninterpreted languages on which a proof theory is definable and a model theory is constructible. Formalizations are always formalizations *of*. In the case of mathematical logic, what is sought for is a formalization of its primary or target properties — entailment, logical truth, consistency, and so on. A formalization is *coherent* to the degree that it has a comprehensive metatheory. A system's metatheory is comprehensive to the extent that it allows for nontrivial consideration of the standard metatheoretical issues. Is the system consistent? Is it sound and complete? Are its target properties decidable? Is its entailment relation compact? Does it have a deduction metatheorem? Is it extrapolative? And so on. A formalization is *mature* to the extent that its answers to these questions are well-developed and rigorous. Coherence and maturity are what one would expect to find in what Joke Meheus calls "decent logics" [Meheus *et al.,* forthcoming].

A useful one-line summary of what has been said so far is that mathematical logic is conceptual formalization, whereas philosophical logic is conceptual analysis. As we have said, one question that the contrast gives rise to is whether a formalization of a property underdetermines its correct conceptual analysis. If it does, a second question arises. Does the underdetermination matter? Throughout the history of logic, there are numbers of cases where it has been asserted that it matters very much, that it amounts to severe conceptual distortion. (As Quine once said about the dialetheic approach to inconsistency, it makes negation "unrecognizable.") A case in point is the dispute between relevant and classical logicians.[3] In classical logic, entailment does not require that antecedents be relevant to consequents. Against this, relevantists argue that, in the absence of a relevancy connection between S and S', it cannot be the case that S entails S'. Here we meet with a question in the philosophy of logic. Naively put, it asks, "What is the true nature of entailment?" Like most philosophical controversies, the present dispute has not been graced with a definitive resolution. However, a kind of holding pattern can be mentioned in passing. It involves a presumed pluralism in logic, according to which "entails" is *ambiguous* as between classical entailment, which first-order logic gets right, and relevant entailment, which first-order logic gets wrong.[4]

---

[3] And modal and intuitionistic logicians, as well as some paraconsistent logicians.

[4] Not to overlook intuitionistic entailment, (non-relevant) paraconsistent entailment, many-valued entailment, dialetheic entailment, and so on.

We can see a certain strategic cunning in the pluralist manoeuvre. It keeps options for the formalization of entailment open as long as the issue remains unresolved. It is as if the mathematical logician is saying, "Look, if entailment is actually classical, then we have a logic for it, namely, classical logic. On the other hand, if entailment is actually relevantist, then we have a logic for it, namely, relevant logic. Until things settle down philosophically, it is best to keep an open mind." As a response to unresolved conceptual conflict, undoubtedly this has a certain attractiveness. But there is also a down-side. The extent to which formalizations can proceed apace and succeed as mathematical constructions in the absence of settled and mature conceptual constraints, they risk promiscuous ambiguation, in which *any* coherent formalization of a notion captures a "sense" of it.[5] Consider, for example, the sheer number of modal logics. They are more or less interesting as mathematical structures (and wonderfully interesting things to study), but it strains belief that corresponding to each different modal system is a different meaning of "possible" in English. So we might have here a case in which, collectively, modal logic *over*determines conceptual adequacy. Of course, there might be many more concepts of possibility than are discernible in the give-and-take of natural language usage. But it is not obvious that there are. This is a matter for the philosophy of language.

It is instructive to take note of two quite different emphases that drive the work of contemporary logic. One is abstract and experimental. It involves building mathematical structures with a view to discovering the logical concepts that can be made to "fall out" of it. It is an adequacy condition on such logics that they have a mature and coherent formalization. In its most extreme form it is *not* a condition on their success that they get even their target concepts "right".[6] "Right" is for philosophers.

The other option *is* philosophical. It takes seriously Eddington's jape that theories are "put-up jobs" (and Quine's that they are "free for the thinking up"). They seek to constrain this latitude by imposing conditions of conceptual adequacy. In actual practice, there is ample evidence that both these forces are in play and that they operate under more or less successful conditions of equilibrium. So in order to get your target concepts *quite* right, it is necessary to get your ancillary or instrumental concepts *somewhat* right; that is to say, right enough. Even so, *in extremis* the two emphases are miles apart.

It is easy to see these contrasts at work in analyses of fictional discourse. Logicians whose interests are primarily mathematical might see a logic of fiction as any mature and coherent system of mathematical logic, in which referring terms and quantifiers for objects of fiction, and truth and closure conditions for fictional sentences are dealt with as part of the logic's overall handling of reference, truth and inference. The mere fact that the theory is mature and coherent would count to some degree in favour of its provisions for its fictional elements. For what is wanted here is not just an *account* of the fictional, but an account that deserves the name of logic. But what really counts depends on what the theory's *primary* targets actually are. If, for example, the logic is chiefly concerned with giving an account of entailment and logical truth, it might well achieve a formalization of those concepts under a number of different conditions, including the following:

---

[5] Left unchecked, we would soon be met with "Anything Goes-Pluralism" (with apologies, in order, to Cole Porter and David Stove [1999].

[6] The impulse to experiment abstractly resembles what drove non-Euclidean geometry before the comforts of an empirically confirmed relativity theory. Here, too, there was no intention of getting space right. That Riemannian geometry did get (big) space right was an inadvertence, and a lucky break for physics.

1. The account it gives of entailment and logical truth is mature and coherent and conceptually adequate.

2. The account it gives of these notions is mature and coherent but conceptually inadequate.

3. The account it gives of entailment and logical truth for fictional sentences is conceptually adequate, but fails to provide a conceptually adequate representation of still other features of fictional discourse, an understanding of which is essential for a philosophically adequate representation of it.[7]

Needless to say, this trio is neither exhaustive nor strictly disjoint.

Perhaps we have said enough provisionally to identify three levels of logical treatment of the fictional.

> *Level 1 (Mathematical).* A mathematical logic of the fictional is a first-level logic to the extent that it is an adaptation of an existing logic without principled regard for the conceptual adequacy of the adaptations. What counts at level 1 is that aspects of the fictional are taken *some* note of and the resultant system is a coherent and mature mathematical logic.
>
> *Level 2 (Conceptual).* A philosophical logic of the fictional is a conceptual analysis of the concept of the fictional. It would stand to a mathematical system as Frege's philosophy of language stands to his formal logic. Although the nature of entailment and logical truth for fictional sentences would be among the theory's targets, there would be an even greater emphasis on reference, quantification and (nonlogical) truth, considered now not as an instrumental necessity but as target properties in their own right. The logic might also engage further parameters of the philosophy of language, including a role for context and user-agency. Accordingly, the theory would take on a pragmatic dimension.
>
> *Level 3 (Conceptually Purpose-Built Formalization).* At this level, a logic of fiction is a mathematical structure whose primary target properties involve all those features of a conceptually adequate account of the fictional, as well as the traditional targets of entailment and logical truth. Such a logic is governed by two kinds of adequacy condition. As a mathematical logic, it must produce a mature and coherent account of its target properties. As a logic sensitive to conceptual nuance, the treatment it affords its further target properties must be philosophically adequate.

Some of the virtues and drawbacks of this trio are readily apparent. Counting for level 1 approaches is their comparative familiarity and their known strengths prior to adaptation. Counting against is the likelihood that their adaptive accounts of the fictional will not be conceptually adequate. Level 2 presents us with all the difficulties attaching to unresolved disputes in philosophy. Level 3 raises questions as to where the limits of mathematical modelling lie. Even so, difficulties and challenges aside, we are now in a position to lay down a corresponding three-step *investigation-procedure* for the logic of fiction. *First step.* Attempt to find an existing mathematical logic which gives a mature and coherent account of its primary properties. Check for conceptual adequacy under fictional adaptation. *Second step.* Whether it is found to be conceptually adequate or not, verify this finding in a conceptual analysis of the fictional. *Third step.* Take what you take to be the best conceptual account of the fictional and try to find a mature and coherent mathematical model for it. To the extent that this proves possible, we will have a mature, coherent and philosophically adequate logic of fiction.

---

[7] For example, are fictional beings incomplete? Are they ever *impossibilia*? What non-deductive inferences are required for a reader to understand a fictional text fully?

Before bringing this section to a close, it would be useful to introduce the idea of *minimal adaptations.*

PROPOSITION 2 (Minimal adaptations). *Let* Σ *be a system of mathematical logic. Let* Σ *be extended by adding to its vocabulary fictional terms and to its grammar fictional sentences. Let* Σ^f *be the result of these changes. Then* Σ^f *is a minimal adaptation of* Σ *if the model theory and proof theory pass to* Σ^f *unchanged except for input constraints.*

COROLLARY 2(a) Clearly, if Σ itself does not count as a logic of fiction, then Σ^f is at best a level 1 logic of fiction.

In a number of ways, the three-step investigation procedure embodies a rather naïve idealism. There is ample evidence that success at step two sometimes inhibits success at step three. It is not difficult to see why. The success of a conceptual analysis of a property might well bring to the fore parameters that resist formalization. They might be conceptually too rich for any extant or foreseeable symbolic representation to digest. Of course, this is an important possibility. It presents us with two options. One is to throw in the towel now. The other is to throw in the towel only after our best efforts have failed. (It is called the principle of best evidence).

2 TARGETS FOR A LOGIC OF FICTION

In considering whether the logic of fiction even has a chance, it is necessary, as I say, to reflect on the *target properties* it is reasonable to suppose that such a logic would seek to elucidate. We turn to that task now.

Systems of mathematical logic routinely set their sights on entailment and logical truth. Should these be considered primary targets of a logic of fiction? Let us begin with entailment. It certainly needs to be determined whether, and to what extent, fictional sentences are deductively closed under the standard logical operations. More basically still, we should want to know something of the conditions under which fictional sentences entail one another, and the conditions, if any, under which a fictional sentence entails a non-fictional sentence. Mainstream systems of mathematical logic also embed certain standing assumptions about *inference.* In a system's proof theory, any deduction of S from $\Sigma = \{S_1, S_2, \ldots, S_n\}$ is taken to licence the inference $\ulcorner$If Σ, then $S\urcorner$. If $S_{1,\ldots,}S_n$ are theorems, then the inference to S itself is mandated. Similarly, if the system's model theory determines that $\ulcorner S_1 \wedge S_2 \wedge, \ldots, \wedge S_n \urcorner$ entails S, the conditional inference $\ulcorner$If $S_1 \wedge S_2 \wedge, \ldots, \wedge S_n, S \urcorner$ is sanctioned. If the S_i are logical truths, then the inference of S itself is sanctioned. No one doubts that this is naïve and rather unrealistic approach to inference. It mandates that there be no extensional difference between the consequence a set of sentences *possesses* and the consequences it is permitted (or required) to *draw* from those sentences. It is easy to see that part of what logic of fiction should try to sort out is the extent to which this view of inference holds water for fictional discourse.

Although entailment and its associated notion of deductive inference make the cut as target properties of a logic of fiction, it should be noted that comparatively little of what makes for a conceptually adequate account of the fictional will show up in the theory's provisions for these notions. The same applies to its logical truths. To the extent that the logic caters for entailment, it will also cater for the associated logical truths. Apart from that, the status of two such truths in particular will require special attention. The law of excluded middle is challenged by accounts in which fictional beings are incomplete, and the law of noncontradiction is challenged by approaches in which fictional entities have logically inconsistent natures or perform logically impossible deeds. Overall, however,

PROPOSITION 3 (The marginality of deduction). *Not a great deal of what is distinctive about fictional entities and their goings-on is captured by the strictly deductive apparatus of their logic.*

COROLLARY 3(a) *In particular, the consequence relations that are most distinctive of fictional discourse are not truth-preserving.*

We saw that in classical approaches to mathematical logic, the concepts of reference and (ordinary, nonlogical) truth are not so much primary targets as instrumental necessities. Classical logic needs *some* kind of story about reference and truth in order to get the main story of entailment and logical truth right, but you can be completely right about entailment and logical truth without having to be completely right about reference and truth. Similarly, you can be completely right about entailment without having to be completely right about deductive inference.[8] As we have said, by the lights of his own philosophical logic, the treatment of reference in Frege's mathematical logic is deficient. But it is adequate, he thought, for the attainment of the mathematical logic's primary targets.

This is very much *not* the case with fiction. Reference and truth are of central importance, both conceptually and formally. In the logic of fiction, reference and truth are not instrumental necessities. They are theoretically primary. It is the same way with inference. The question of what can be inferred from a literary text is fundamental to what [Parsons, 1980] calls the "maximal account" of it, in relation to which the written text is a kind of extended enthymeme.

Doubtless there will be features of a comprehensive conceptual account of fiction, e.g., matters of style or the literary significance of irony, that exceed the reach of anything that could reasonably be called a logic. However,

---

[8] Similarly, a theory might produce a wholly credible analysis of entailment and logical truth, based in part on counterintuitive treatments of, say, negation (in, e.g., a logic of first-degree entailment).

PROPOSITION 4 (Target properties). *Since entailment, logical truth, reference, (ordinary) truth and inference are squarely within any logic's ambit, we shall, as a first pass, make these the target properties of a logic of fiction.*

3 INTUITIONS

The targets proposed by Proposition 4 are well-supported by the more basic of our pretheoretic intuitions about literary fictions. The following five appear to be centrally important [Woods, 1974]:

A. Reference is possible to fictional beings even though they do not exist.

B. Some sentences about fictional beings and events are true.

C. Some inferences about fictional beings and events are correct.

D. These three facts are made possible in a central way by virtue of the creative authority of the authors of fiction. Indeed, the primary and originating criterion of truth for fictional sentences is the author's sayso.

E. It is also possible for a fictional truth to make reference to real things. For example, "Sherlock Holmes lived in London" is true and refers to the actual capital city of England.

If we wished to exploit an analogy with the axioms of naive set theory, we could speak of these as axioms of intuitive theories of fictionality or the *intuitive axioms for fiction*.

No one should think for long that the intuitive axioms are problem-free. It is true that axiom *A* appears to take on a degree of corroboration from the structure of quantification in natural languages. In classical logic, "There are things that don't exist" is an inconsistent (or in some versions a non-well-formed) sentence. In English, it is well-formed, consistent and *true*. Nevertheless, philosophers have long struggled to produce accounts of how we manage to achieve reference to existing objects. As I write, the leading such theory — the causal theory — cannot be extended (or so it would appear)[9] to putative references to the nonexistent. If the causal theory of reference gives our best account of reference, it is disappointing that reference to the nonexistent lacks as good a theory, if any at all.

Perhaps the biggest problem that besets axiom *B* is that for well over a half-century of tightly focused attention by philosophers of language and model-theoretic logicians, truth is tied to reference. So if reference to the fictional cannot be reconciled to our best theory of reference (if any), neither can truth about the fictional be handled by our best theory of truth (if any). A second problem for *B* can be loosely described as follows. Let S be any sentence from a story. Then the world is such that S is not true. In some cases, the contradiction is more direct. For example, *B* allows us to say that we know that Holmes lived in London. But what we know of London allows (indeed requires) us to say that he did not. One possibility is that these are only apparent contradictions, that they disappear once the ambiguity of such sentences is duly noted. But *where is the ambiguity*? The sentence, "Holmes lived in London", gives no sign of lexical or syntactic ambiguity. Perhaps it is an ellipsis of some kind. But if so, care needs to be taken in unpacking it lest unnecessary damage be done to the axioms.[10]

---

[9] As witness e.g. [Kripke, 1972; Kaplan, 1973] and [Plantinga, 1974].

Axiom C tells us that sound and accessible inferences exist from sets of fictional sentences. But if there are difficulties in deciphering what, if anything, they are about, and linked difficulties in making out the conditions, if any, under which they could be true, it is hardly surprising that similar uncertainties would afflict the matter of what inference rules are correct for fiction. For example, if 'a' is "Prime Minister Blair" and 'b' is "Sherlock Holmes", then

$$\begin{array}{rl} & Fa \\ \therefore & \exists x Fx \end{array}$$

is a classically good inference, whereas

$$\begin{array}{rl} & Fb \\ \therefore & \exists x Fx \end{array}$$

is problematic. Of course, in this case the variable of quantification may range over the nonexistent as well as the existent. But this subjects inference to the same uncertainties as attach to A.

Philosophers of language are familiar with Austin's notion of a *performative utterance*. It is widely agreed, for example, that in uttering, "I promise to wash the car" I have, in the absence of particular reasons to the contrary, created a promise. I haven't just said that I promise; I've actually promised. Performative utterances have attracted some successful attention in theories of speech acts [Austin, 1975; Searle, 1970; Grice, 1989]. There would be a clear advantage if the creative aspect of an author's sayso could be modeled in a speech-act pragmatics. The trouble is that the statements by virtue of which fictional objects and events are created seem not to have the character of performatives.[11]

A further puzzle is whether, and how, an author's sayso can make contradictions true. Even if were able to deal with the kinds of contradictions mentioned above, this would not help with the problem of directly authored contradictions such as the case in Ray Bradbury's *A Sound of Thunder* in which Keith both is and is not elected president in one and the same election in 2055. One might think that such sentences could be accommodated in a dialetheic logic. Dialetheic logics aren't particularly congenial to the providence of axiom A, but perhaps this might be fixed.

3.1 Pretense and make-believe

To date, the logic of fiction research programme has concentrated on literary fictions, as opposed to legal and mathematical fictions, to name just two. Important as this focus assuredly is, there is a potential drawback. It lends encouragement to the idea that a logic of literary fictions is somehow basic, or canonical, for fictions of all types; in other words, it appears to assume *The Literary Primacy Thesis*. Upon reflection, however, it is worth pausing to consider whether we might be better served by thinking the opposite, namely, that

---

[10] See here [Woods, 1974, chapter 2].

[11] This is disputed in [Åqvist, 2003].

PROPOSITION 5 (Fiction as generic). *It might turn out to be the case that there is a generic logic of fiction of which the logic of literary fictions is a special case.*[12]

I shall not at this stage attempt to decide the correctness of Proposition 5, or of *The Literary Primacy Thesis*. For the time being I shall concentrate on theories of literary fictions, returning to Proposition 5 towards the end of the chapter.

Let us now supplement our intuitive axioms with some further commonplaces. A fictional fact is not a real fact. A fictional person is no person at all. No fictional event ever occurred. Fictional properties don't exist. These and other like features of the fictional are a natural setting for an intuition of broad appeal.

PROPOSITION 6 (The Pretense/Make-Believe Model). *Fictional objects and facts are the subjects (or contents) of pretense.*

COROLLARY 6(a) Discourse that is nominally about fictional objects and facts is a form of make believe.[13]

The tug of the pretense/make-believe (*PMB*)-intuition draws emphatic attention to the intuitive fact that when it comes to fictional entities *there is nothing there*. It lays corresponding emphasis on the intuitive fact that fictional discourse *leaves no metaphysical footprints*.

A difficulty with the *PMB* approach is that it suggests a certain general playfulness. Doubtless some pretendings — children in a game of Cops and Robbers for example — are the sheerest play. How close the *PMB*-intuition drives the philosophical logic toward these paradigms is something we shall have to consider with due care. This task is taken up in §4.1.

3.2 Can't help it realism and the impulse to reify

Against the *PMB*-intuition something equally strong countervails. For lack of a better name, we'll call it the *objectual intuition* (the *O*-intuition). It embeds the idea of axiom *A*:

PROPOSITION 7 (The O-model). *In some nontrivial sense fictional beings are indeed objects and fictional goings on are indeed events.*[14]

Underwriting the *O*-intuition is the fact (or apparent fact) that unlike terms such as "the present king of France", which have no denotations at all, terms such as "Sherlock Holmes" are genuinely denoting, albeit not of existent objects. Accordingly, as [Woods, 1974] points out, statements about Sherlock Holmes, unlike those "about" the present king of France, are *bet-sensitive*. There is a bet to be won or lost as to Holmes' place of residence. "221B Baker Street" wins. "2237 W. 14th Avenue" loses. Concerning the present king of France all such bets are lost (vacuously).[15] [16]

---

[12] This is the position adopted by [Woods and Alward, 2002].

[13] An early example of authorial pretense analyses is [Searle, 1975]. See also [Scruton, 1974]. The leading reader-appreciator analysis is [Walton, 1990]. [Lewis, 1983] is also a pretense theory.

[14] *O*-approaches include [Kripke, 1973; Routley, 1980; Woltersdorff, 1980; Lewis, 1983; Zalta, 1983; Deutsch, 1985; Lamarque and Olsen, 1994; Thomasson, 1998; van Inwagen, 2000] and [Zalta, 2000].

The *O*-intuition seems to be grounded in the kind of cognitive being the human animal is. The human animal is a "can't-help-it realist."

PROPOSITION 8 (Can't-help-it realism). *Whether by nature or by nurture, the human cognitive agent is rigged so as to take the realist stance, that is, to experience the world realistically.*

Perhaps it all has to do with survival, since the world is dangerous and realism is an efficient way of paying attention. In our Kantian moments, we might think that experiencing the world realistically is a condition of there being experiences at all. Smart skeptics (of whom Hume was one and Sextus not) fully well know that scepticism is not a livable truth if indeed true, that one could no more experience the world as acausal, or as wholly interior, or as mindless except for me, than jump over the moon.

PROPOSITION 9 (Smart scepticism). *Smart sceptics, therefore, allow that even though they may concede that realism is false, the human animal is a can't-help-it realist.*

Our can't-help-it realism gives us natural occasion to reify. When seized by an onrushing tiger experience, we take it that a tiger is rushing on. Of course, this is no ordinary postulation. It happens automatically and in a blink. Reifications elsewhere take longer. A case in point is the reification of *ZF* sets. When in 1902 Russell conveyed to Frege the news that killed intuitive set theory dead, a philosophical crisis ensued. The critical question was whether set theory could be rehabilitated and, if so, under what conditions. Russell answers this question in *The Principles of Mathematics* [Russell, 1903]. In one way, says Russell, set theory could not be rehabilitated. The paradox that bears his name rendered it impossible to generate the axioms of sets by a *philosophical analysis* of the idea of set. On the other hand, Russell supposes that a kind of rehabilitation could be achieved by what he called *mathematical definition*, which he sees as a form of nominal definition. In other words, the project of revealing the true natures of what sets actually are was pole-axed by the Russell contradiction, but it remained possible to re-introduce sets by *stipulation*. Russell was fully aware that stipulated objects are not free for the asking. Stipulations were also invitations to use. Their mathematical legitimacy would be decided by the mathematical research community. Russell's own contributions to this competition fell by the wayside, and in time *ZF* became the theory of choice (actually, *ZFC* if the pun might be forgiven). *ZF* arose in the axiomatic stipulations of Zermelo as early as 1908, but with the passing decades it came to be regarded as the authoritatively correct description of the intuitive idea of set. It was a remarkable transition, from fiction to fact. It was occasioned by impulse to reify that flows unstoppably from our can't-help-it realism.

---

[15] Consider a case. If the present king of France is the king of France, then where does he live? Does he live in Paris or Versailles? Or possibly in Avignon? Perhaps he's in exile in Atlantic City? These are transparently silly questions. All bets are off.

[16] In [Woods, 1974] the contrast between bet-sensitive and bet-insensitive terms is also reflected in the contrast between *nonesuches* and *non-existents*. Whereas there is no such thing as the present king of France, there is such a thing as Sherlock Holmes albeit not a thing that exists. The present king of France is a nobody. Holmes is a somebody, albeit a nonexistent one. For doubts about these contrasts see [Griffin, 2005].

Can't-help-it realism doesn't settle the question of whether realism is true. It calls to mind the distinction between strong and weak AI. According to strong AI computers (or computer programs) have minds in just the way that humans do. Weak AI allows that strong AI may be false, but holds that the study of mind is methodologically advantaged by the assumption that it is true. We might draw a similar distinction between strong and weak realism. Our own can't-help-it realism is pairwise compatible with each. But it is *like* weak realism. It suggests that, all things considered, we get a better appreciation of how things are by experiencing the world realistically and of yielding to the impulse to reify.

3.3 Modalities de dicto *and* de re

It is natutal to think of the tug between *PBM* and *O*-models as paralleling that between *de dicto* and *de re* modalities. This has something to do with the interplay between quantifiers and the syntactic devices that betoken the fictional. One of the founding data of most accounts of fiction is that, while it is true that none of it happened, there are nevertheless well-individuated entities of which it is true. The *PBM* and *O*-orientations incline their respective supporters to different formulations of this fundamental insight. *PMB*ists see it on the model of

1. It is true in the story that $\exists x(\ldots x \ldots)$.

Those of the *O*-persuasion are more drawn to

2. $\exists x$ (it is true in the story that $\ldots x \ldots$).

If we put T^F as short for "it is true in story S that", the *PMB*-formulation becomes

3. $T^F \exists x(\ldots x \ldots)$

whereas the *O*-formulation comes out as

4. $\exists x T^F(\ldots x \ldots)$.

If we also followed the practice of a good many theorists of fiction in regarding T^F as a kind of *modal* operator on sentences, then the *de dicto*/*de re* parallelism is quite clear. In (3) the fictive modality is *de dicto*, having no occurrence within the scope of the quantifier. In (4) the modality is *de re*; it occurs in the scope of x. It is a virtue of (3) that its truth leaves it undisturbed that the fictional leave no metaphysical footprints.[17] One of its drawbacks is that it trifles with the intuition that there are well-individuated objects of which the goings on in fiction are true, as well as the kindred intuitions that they are also subject to truths emanating from us rather than from the story.[18]

---

[17] As Kendall Walton once remarked to me in conversation, it is of central importance to a correct account of fiction that, while there may be objects of fiction, no *object* is an object of fiction, or could be.

[18] Richard Routley, also in conversation, once pointed out that, given that Agatha Christie admired Sherlock Holmes, not only was there some object of which that is true, but that it is something made true by Christie, not by Conan Doyle.

On the face of it, the *O*-interpretation is more accommodating. That Holmes is a well-individuated object of which certain things are true is indicated by the quantifier of (4) in *de re* position. That the things that are true of it (or him) never happened — to him or anyone — is flagged by the fictive operator on the scope of (4)'s quantifier.[19] It also bears on the advantages and disadvantages of (3) and (4) that it is much more natural for us denizens of the world to speak of Holmes and his ilk, and on their to-ing and fro-ing, in the manner of (4); whereas (3) is held in reserve for when (4) is under *philosophical* challenge of the sort pressed by no-object theorists such as Walton. In natural speech under normal circumstances, quantifiers show a marked tendency to poke to the left. Doing so gives syntactic expression to our fondness for reification. Our leanings towards reification find a safe harbour in attributions *de* re. Reification is *de re*-ification. We might say that in speech and thought we are predisposed to handle quantification in modal contexts in conformity to the Barcan formula. So conceived, quantifiers have a preference for the left. As we are presently characterizing the presumed tension between (4) and (3), if we found reason to doubt (4), then (3) would be the only alternative. We should reconsider this. A further attraction of the *O*-model is the promise it holds out for reconciling the two, roughly as follows:

5. For any interpretation under which (4) is true, (3) is also true.

6. There are interpretations under which (4) is true.

An interpretation under which both (4) and (3) are true is one in which, for the values of the quantifier in (4) objecthood is assured but actuality is not. I shall return to this point in §§4.2 and 4.3.

The transition of *ZF* sets from fictions to facts, reminds us of the necessity to recognize a third intuition about fictions. This is reflected in the S-model.

PROPOSITION 10 (The S-model). *Fictional objects arise from stipulations and, in the first instance, fictional truths are made true by stipulation.*

The S-model brings into consideration a certain way of making of our can't-help-it realism (and the embedded disposition to reify) not only states of mind forced upon us by nature and circumstance, but states of mind having genuinely objectual content. This is precisely what they would be were it the case that stipulations that meet certain minimal constraints are metaphysically *creative*. This was Hilbert's view of mathematics, for whom any consistent stipulation of a mathematical entity produced that entity, brought it into being. Of course, there is something weaker to say about creative stipulations. It is that stipulations aren't creative in fact, but that our can't-help-it realism induces us to see them as such. So there is an important question to be considered. Do stipulated beings leave metaphysical footprints or are they metaphysically more ephemeral than that? Here too we see the *de dicto*/*de re* distinction. In purely instrumental versions, stipulations verify sentences in the form $\ulcorner T^F \exists x(\ldots x \ldots)\urcorner$ but not in the form $\ulcorner \exists x T^F(\ldots x \ldots)\urcorner$. In objectual versions it is the other way round.

---

[19] (4) does not however seem to capture Christie's admiration of Holmes, in as much as her admiration is a fact of the world, not of fiction, and seems not to belong in the scope of the fictive operator. I return to this question in §5.4.

This would be a good place to make the present point more generally and to note that

PROPOSITION 11 (A slack trichotomy). *Our three models are not all on a par, and are not entirely disjoint.*

It is easy to see why the *PMB*-model leaves no metaphysical footprints. But the same might also be said of some versions of the *O*-model. If so, the model could postulate genuine objects that nevertheless "aren't there." On the other hand, some versions of the *S*-model have it that stipulated objects are indeed there (it's just that they weren't there prior to stipulation.) Similarly, we might find theories in which the objects are stipulated yet their subsequent adventures are a matter of make-believe. If this is right, the *PMB*-model is itself a hybrid. So there is some slack in the space occupied by the *PMB*, *O* and *S*-models.

In the large literature that has grown up around these issues, perhaps [Walton, 1990] is the purest expression of the view that there are no fictional objects in any sense of "are" or "object", as are [Parsons, 1980; Jacquette, 1996] and [Griffin, 2005] the purest expression of a strong realism towards the fictional.[20] Virtually everything else lies somewhere between these extremes, with the majority of these intermediate positions clustering around a loosely structured notion of weak realism.

3.4 Fictional discourse

It is necessary to distinguish three distinct sort of fictional discourse, all of which require elucidation in a conceptually sensitive logic of fiction. Following [Currie, 1990] we will say that *fictive* discourse consist of those sentences which constitute fictional texts [Doyle: 1981]:

- Holmes was sitting with his back to me, and I had given him no sign of my occupation (p. 7).
- 'Good!' said Holmes. 'Excellent!' (p. 7).
- Sherlock Holmes waved our strange visitor into a chair (p. 12).
- Holmes stretched out his hand for the manuscript and flattened it upon his knee (p. 14).

Metafictive discourse consists of sentences that are not fictive, but which may be inferred or surmised from literary texts.

- Holmes was a detective
- Holmes lived in London
- Holmes solved the Case of the Speckled Band
- Holmes patronized Watson

---

[20] See also [Parsons, 1975], [Routley, 1979; Castãnada, 1979] and [Zalta, 1983].

- Holmes was a marvellous reasoner.

Such sentences appear to make claims true of the stories, yet false of the world, to speak loosely.[21] For example, it appears to be true of Holmes that he lived in London but not true of London that Holmes lived there [Woods, 1969].

Transfictive discourse consists of sentences are also used to make claims about the characters and events that occur in works of fiction, but they make particular claims about their relations to things external to the works in which they occur.

- Holmes was admired by Agatha Christie
- Holmes was appropriated in a later novel, *A Monstrous Regiment of Women*, by Laurie R. King
- Jeremy Brett caught nuances of Holmes' personality better than any other actor to date

Such sentences might be said to make claims that are true of the world, yet false of the stories, still speaking loosely.[22]

A principal part of the logician's task is to set forth the production rules of stories' maximal accounts. Here the concept of presumption plays a key role. There is a widely held intuition that the full story of, for example, *The Hound of the Baskervilles* is not exhausted by the totality of the lines constituting Conan Doyle's printed text, that is to say, by the fictive sentences to be found there. There is much there that it is left to the reader to "fill in" by way of metafictive sentences. This, too, is a kind of closure exercise, but it is closure under weaker consequence relations than that of deductive implication. So a further requirement of the logic of fiction is to specify these consequence relations and determine their closure-conditions. A case in point: It is nowhere recorded that Sherlock Holmes had an aliamentary canal. Should we presume, metafictively, that he did have one? If so, on what basis? Do we repose our trust in the generalization that aliamentary canals are standard equipment for people? Or should we allow that since fictional people *aren't* people, what characteristically holds of people may well not hold of Holmes? In that case, should we try to locate Holmes' aliamentary canal in the generalization that all fictional people have them? The trouble is that it is precisely *this* that is in question here.

Consider a second case. With regard to transfictive or mixed sentences, such as "London was the city of Holmes' residence", the quite general question arises as to what London (and the world, for that matter) would have to have been like for that sentence to be true. For example, do Holmes' city and the world in which it resides obey the laws of physics? Varying the example, it happens Cervantes' *Don Quixote* was published in two parts, one in 1605 and the other in 1615. Kepler's first two laws of planetary motion were published in 1609; the third in 1619. Does this mean that the heavenly bodies under which Don Quixote slept in the first part described circular orbits, yet those of the second part described elliptical orbits? [Woods, 1974, 65].

---

[21] Of course, a sentence of metafictional discourse containing the name of a real person might turn out be true of the world as well.

[22] A sentence of transfictive discourse might also turn out to be true of the story.

3.5 Difficulties with pre-existing accounts

3.5.1 Descriptivism

Many philosophers are of the view that Russell's motivation in developing the theory of descriptions [Russell, 1905] was as a means of handling the problem of negative existentials [Cartwright, 1960]. If this is right, it is natural to look to Russell for instruction in matters fictional, since one of the problems thrown up by fiction is the nonexistence of its objects. How, then, are we to express this negatively existential fact? As developed by Russell and extended to names by Quine [1939], sentences about Holmes are false, just as sentences "about" the present king of France are false, and the intuitive difference between these referring expressions is wholly erased. In that regard, descriptivism won't do at all for fiction. It might also be remarked in passing that, although Russell was interested in the problem of negative existentials, this was not in fact the motivation for "On Denoting". What Russell wanted the theory of descriptions for was to elucidate the contrast between knowledge by acquaintance and knowledge by description. This is also a question of direct salience for fiction. Given that we cannot be acquainted with him, how is our knowledge of Holmes structured? "On Denoting" embeds a relevant question for fiction, but gives the wrong answer. Accordingly,

PROPOSITION 12 (Minimal adaptations of the theory of descriptions). *Since they are minimal adaptations of conceptually inadequate logics, description logics are at best level 1 logics for fiction.*

3.5.2 Free logics

A free logic in the narrow sense is a quantificational system whose domain of interpretation can be empty and whose individual terms need not denote (but when they do they denote actual objects). Sentences containing nondenoting terms are classically truthvalueless, thus violating axioms A and B. A narrowly free e-logic is a free logic with a dedicated existence predicate [Lambert, 1963; 1963a; 1964] and [van Fraassen, 1966; 1966a].[23] Such logic fails to preserve the contrast between terms for nonesuches, such as "the present king of France", and terms for somebodies that don't exist, such as "Sherlock Holmes" So these logics violate axiom A.

Free logics were purpose-built for nondenoting terms. The question here is whether they can be adapted to include fictional terms as a special case. Such a view of fiction is conceptually inadequate, never mind its formal virtues. Accordingly,

PROPOSITION 13 (Minimal adaptations of free logics). *Since they are minimal adaptations of conceptually inadequate logics, free logics are at best level 1 logics for fiction.*

Let us also add that a logic is a free logic in the broad sense if it admits nondenoting terms without paraphrase or whose terms may denote non-actual individuals. (An exception is Lejewski [Lejewski, 1954], which — rightly in my view — doesn't tolerate empty singular terms such as "the present king of France".) Routley's existence-neutral quantification theory [Routley, 1966], is a broadly free e-logic. But such logics cannot accommodate mixed sentences such as "Holmes lived in London"; hence they disconform to axiom E.

---

[23] [Hailperin and Leblanc, 1959] is also a narrowly free e-logic. [Hintikka, 1959] is not, despite some similarities. [Smiley, 1960] too resembles, without quite being, a narrowly free logic.

3.5.3 Supervaluational logics

A supervaluation logic is a logic in which there is a way of assigning truth values to sentences of a formal language which allows that some sentences are classically nonbivalent and yet none of the standard theorems is lost [van Fraassen, 1966; Bencivenga, 1980; 1991]. But here too the intuitive contrast between "the present king of France" and "Sherlock Holmes" collapses.[24] However, non-bivalence approaches to fiction persist. See, e.g., [Taylor, 2000; Adams *et al.*, 1997], and [Slater, 2002].

PROPOSITION 14 (Minimal adaptions of supervaluational logics). *For reasons cited above, supervaluational logics are at best level 1 logics for fiction.*

3.5.4 Epsilon calculi

Of note here are [Slater, 1987] and [Slater, 2002], in which it is proposed that the inducements of many-valued and supervaluational approaches to the logic of fiction can be obviated by way of epsilon terms. In an epsilon calculus — which is a conservative extension of the predicate calculus — there corresponds to every predicate 'F' the epsilon term 'εxFx', which denotes a given F provided there are some Fs, and "has an arbitrary reference otherwise" [Slater, 2002, 181]. Epsilon calculi have the resources to explain how it can be the case that something is the one and only F even though there is no such thing. As Slater observes, this is a feature that makes the epsilon approaches an attractive way of handling fictional discourse. For although there is no such person as Holmes, it was he and only he who solved *The Case of the Speckled Band.* Even given that Holmes is the one and only person who is Holmes, the epsilon calculus offers us the good advice not to try to book an appointment with him, yet without having to give up on the question of who solved *The Case of the Speckled Band.* A crucial test, as always, is whether the epsilon approach preserves the distinction between the present king of France and Sherlock Holmes. The answer is that it doesn't and it does. At the level of reference, "the present king of France" and "the solver of *The Case of the Speckled Band*" are at par. At the level of truth things are different. The truth of e.g., Holmes' residency is conveyed by a map from {"Holmes lived in Baker Street", i} to the truth-value 1. Here 'i' denotes the stories that make this true. As for the residency of the present king of France, there is nothing to be known; this is because there is no authorizing i for the present king of France. Some readers will think that the requisite contrast is not deeply enough drawn here. If we have it that he who presently reigns in France is the one and only who does, and yet that there exists no one who presently reigns in France, and that he who solved *The Case of the Speckled Band* is the one and only person who did, and yet there exists no one who solved *The Case of the Speckled Band*, the difference between them is that there are stories about the latter, and none so far about the former. As with Meinongian approaches, the subjects of fictions precede their fictionalization. This is tantamount to a violation of axiom D — that fictional beings originate with their ficationalizations. The distinction between "Sherlock Holmes" and "The present king of France" is also lost here. Accordingly

---

[24] We note in passing that neither free logic nor supervaluation semantics has had the legs intended for them if we are to judge by, e.g., [Everett and Hofweber, 2000].

PROPOSITION 15 (Minimal adaptations of epsilon calculi). *These too are at best level 1 logics for fiction.*

3.5.5 Meinongian logics

Meinongianism also runs afoul of axiom D — the author's sayso axiom. On this view, all objects, fictional or otherwise, *are* by virtue of their corresponding sets of properties. Thus fictional objects "were objects before they were written about; they were so to speak only identified, or perhaps discovered, by the author, and writing about them did not confer objecthood upon them" [Parsons, 1975, 79]. The requirement that fictional objects be run-of-the-mill Meinongian objects precludes it being the case that in a rather deep and somewhat literal sense fictional objects have a literary paternity; i.e., that they are created by their authors.[25]

Particular note should be taken of Jacquette's [1996]. It is a good example of a mature and coherent mathematical logic. It has a well-specified formal language, and the requisite proof and model theories. The system is sound and complete. It has deduction, compactness and interpolation theorems. It is a three-valued logic in the manner of Łukasiewicz. It is undoubtedly a "decent" logic in Meheus' sense. It is inspired by a conceptually mature metaphysical theory. It was purpose-built to accommodate Meinong's theory of objects, which notwithstanding the slurs of Russell is an intellectually robust rival of the timorous ontologies of classical logic. Of special note is its satisfaction of axiom A. In Jacquette's hands, the logic of objects adapts in an apparently natural way to the logic of fiction. It is a fair approximation of a level 3 logic of *Meinonge*. And while it handles fictional language with all the formal adeptness it affords to Meinongian language, it cannot be regarded as a level 3 logic of fiction. It is not adequate for the conceptual demands of fiction. Its adaptation to the demands of fiction is minimal. I recur to this point in greater detail in §4.2 to follow.

3.5.6 Theories of partial objects

Theories of partial objects [Landman, 1986] permit the introduction of formal objects with out identity criteria in *ATMS models* that are designed to consider hypothetical objects in the context of alternative abductions of bodies of evidence or data [de Kleer, 1986]. Informally, the basic idea is that in a criminal investigation, various possible theories of the case are often considered. In an incriminating theory, an account of the findings is predicated of some determinate individual. In a non-incriminating theory, an account of the case is predicated of some merely possible object. The theory of partial objects develops a semantics for the sentences of such accounts that purport to refer to such "objects".

---

[25] [Kivy, 1993] argues that musical works are discovered and not created. Presumably similar arguments could be mobilized on behalf of the thesis that fictional characters pre-date authorial creative activity. Levinson, however, has offered reasons for balking at such arguments [Levinson, 1980].

The partiality of hypothetical objects squares well with any theory of fiction in which fictional objects are incomplete. Even so, lost in the partial objects approach is the intuition that fictional beings, whether incomplete or not, are well-individuated by their maximal accounts. So, here too, partial objects logic of fiction could at best be only a level 1 logic.

3.5.7 Paraconsistent logics

The principal feature of paraconsistent logics is that they are contradiction-tolerant [da Costa, 1998; Schotch and Jennings, 1989; Brown, 2000; Batens, 2000; Meheus *et al*, forthcoming; Jacquette, 1998]. Paraconsistent consequence does not satisfy the classical provision that a contradiction implies every statement. To the extent that axiom D — and also axiom E — allow for intended contradictions to hold in a work of fiction, it is certainly arguable that deductive closure in fictional discourse is subject to paraconsistent constraints. However, as they stand, such logics make no provision for axioms A, or for B, C, D or E. Accordingly, any successful minimal adaptation would be at most a level 1 logic of fiction.

Here, too, we meet with strongly competing theoretical intuitions. Meinongians such as Jacquette are at ease with the idea of nontrivial reasoning from inconsistent data-bases, which they effect with the device of predicate complementation. Theorists, such as Lewis, who abjure the very idea of "allowable" inconsistencies, preclude nontrivial inferences from inconsistencies by imposing severe constraints on fictional reasoning as such. It falls to the logician of fictional to adjudicate these rivalries. Lewisen insights are further explored in §4.1.

3.5.8 Dialetheic logics

Dialetheic logics are paraconsistent, with a difference. Whereas paraconsistent systems are contradiction-tolerant, dialetheic logics are contradiction-*affirming* [Routley and Meyer, 1976; Priest, 1979; 1987; Armour-Garb; 2005; Woods, 2005]. If we persist in the view that D and E have the force of verifying fictional contradictions, deductive closure in fiction would have to be dialetheic, except as otherwise provided by their author's fictive sentences. But notwithstanding [Heintz, 1979], dialetheic logics are not, as they stand, especially congenial to fictional objects and fictional truths. Consequently, here too, minimal adaptation would at best provide a level 1 logic of fiction.

With the dialetheic turn, we appear to lose the sympathies of virtually anyone who has attempted an analysis of fictional discourse. In fact, it is a dispute in which dialetheists have a certain tactical advantage. Why, they will ask, if it is inferentially harmless to admit contradictions to literary texts, what *further* harm would flow from their sometimes being true? All paraconsistentists agree that, whatever else, it should not be allowed that from an inconsistency everything be inferred. But apparently it is possible to avoid the disaster of *ex falso quodlibet* without forbidding the very idea of true contradictions.[26] That being so, it remains an open question in the logic of fiction as to whether a fictional truth might be a contradiction.

3.5.9 Substitutional quantification

We come now to an early attempt to move beyond level 1. Sentences containing quantifiers are usually interpreted by assigning a value to a contained variable. Values are construed as objects from the theory's domain of discourse. However, when considered substitutionally [Woods, 1974], *terms* (not objects) are assigned to quantified variables.[27] By these lights, "Someone solved the Case of the Speckled Band" would be true just in case, for some name N, "*N* solved the Case of the Speckled Band" would also be a true sentence. If *N* were the name of Sherlock Holmes, the sentence would indeed be true. One of the attractions of the substitutional approach to fiction is that it offers an engaging solution to the problem of negative existentials; it frees quantifiers from the burden of existential import. Accordingly, "There are things that don't exist" could turn out true on the strength of the truth that Sherlock doesn't exist, without the embarrassment of the embedded name "Sherlock"'s having to denote an existent object. But there are problems. One is to find the right link between the reference of "Sherlock" and the truth of "Sherlock solved the Case of the Speckled Band". As mentioned above, it would appear that these objectives can't be met in the standard (and best to date) theories of reference and truth. An alternative might be to fashion a theory of truth in the manner called for by *C*, and leave fictional names without referents. Doing so, however, would violate axiom *A*.

[Woods 1974] is purpose-built for fiction. It is a nonstandard modal extension of classical logic. Apart from the trouble occasioned by its failure to fit axiom *A*, additional problems threaten. One is that it might fail to be mature and coherent mathematical logic. This is, in effect, precisely what the criticisms of [Howell, 1976; Parsons, 1978] and [Routley, 1979] claim. A further difficulty might be that the conceptual model that [Woods 1974] takes pains to articulate is underdetermined by its own mathematical logic. Such indeed is the case. For example, Woods' formal logic makes no provision for inference relations peculiar to the construction of maximal accounts.

3.5.10 A methodological reflection

As our brief examination of pre-existing logics makes clear, there is a natural tendency among investigators to assimilate their findings to what is already known. We see in this the guiding presence of the *Can Do Principle*. *Can Do* helps regulate the cognitive economies in which scientific and other forms of enquiry occur. It is a principle of methodological conservativism, bidding the researcher to transact his business with tools he already commands and to repose his results in theories that are well-understood and have stood the test of time. At bottom, *Can Do* is stout discouragement of re-inventions of the wheel. It is a motivating force behind reductionism in most of its varieties and it is a presupposition of many attempts to unify knowledge. Like many good principles, *Can Do* has a degenerate case. At its most extreme, *Make Do* is *Can Do* run amok. At its least extreme, *Make Do* abuses the perfectly reasonable principle that often it is better to have an imperfect theory than no theory at all. In the hands of *Make*-Do, *wrong-headed* theories tend also to make the cut. *Make Do* inclines the theorist to employ tools that don't fit his present task and to seek a home for his results in theories that cannot justifiably accommodate them, on the grounds that the appropriated tools and borrowed theories are well-understood and have scored impressive successes in their rightful spheres of influence. Accounts fashioned in this way suffer from what [Toulmin, 1953] calls *procrusteanism*.

---

[26] This issue is explored in greater detail in [Woods, 2005].

[27] [Pavel, 1986; Doležel, 1979] and [Currie, 1990].

It is an abiding difficulty of theoretical analysis that degenerate cases of *Can Do* are so often not self-announcing and that they make themselves known only after the fact, if at all. Similarly, some of the most fruitful reductions, as with Descartes' algebraic reworking of plane geometry, have been considered wildly implausible before the fact. In general, we do not have a particularly reliable working command of the *Can Do – Make Do* distinction. So one does the best one can and waits to see what happens. The logician's desire to get a handle on the fictional is wistful occasion to bear these difficulties in mind. Like it or not, he is caught in the tug of two forces. To the extent possible, he should assimilate the account of fiction to what he already knows how to handle. To the extent possible, he should also avoid the excesses of *Make Do*.[28] Readers sympathetic to this line will likely agree with Penelope Maddy on a similar issue.

> ... philosophers seem occupationally disposed to analogies: mathematics is like a game, or mathematics is like fictional story-telling, or mathematical language is like metaphorical language ... Let us imagine that [a philosopher of mathematics] instead undertakes to characterize mathematics *directly, as itself*; instead of trying to understand mathematics by analogy with something more familiar, she tries to make mathematics *itself* more familiar. [Maddy, 2005, 364], emphases added).[29]

As we see, the usual run of pre-existing logics seem not to work convincingly for fiction. Perhaps this was only to be expected. Perhaps a happier outcome will be achieved by a logic that has been wrought with the imperatives of a conceptually adequate appreciation of the fictional firmly in mind. In the section to follow, we shall examine in some detail the dominant conceptual models of the fictional. Of the three, as we shall see, only the S-model evades the depredations of *Make-Do* although, even there, not everything is clear sailing. So we shall now flag a negative thesis.

---

[28] *Can Do* and *Make Do* are discussed at greater length in [Woods, 2006]. But we might here note further variations. As we have it so far, *Can Do* and *Make Do* are about the appropriation of existing *theories* in an effort to bring the analytical task at hand to heel. The same processes are also at work with regard to one's choice of an *analytic vocabulary* for a problem. Two quick examples will give the flavour of this. First, it is common among modal semanticists to analyse possible worlds as maximal consistent sets of propositions. Modal sceptics such as Quine complain that the analyzing idiom of "proposition" is no better-understood than the *analysandum* "possible world". Secondly, it is still widely accepted to define "analytic sentence" as one that converts to a logical truth upon the replacement of terms by synonyms. Intensionality sceptics (Quine again) complain that the analyzing idiom 'synonym' is no better-understood than the *analysandum* 'analytic'. In each case, the sceptic is pressing a version of the *Make Do* complaint. (I cite these examples without prejudice.)

[29] Cf. Michael Oakeshott on what the poet does: he imagines *poetically*.

PROPOSITION 16 (PMB, O and Make-Do). *The dominant versions of the PMB-model and the O-model suffer, in varying degrees, from the debilitations of Make-Do.*

4 EXAMINING THE BASIC CONCEPTUAL MODELS

The upshot of the forgoing discussion is that existing logics have not responded well under minimal adaptation to the requirements of a logic of fiction. Even logics such as [Woods, 1974], which was specifically constructed for fiction, fail to formalize a number of conceptually significant features. In effect, it too is a minimal adaptation. In the present section, we turn our attention to the various ways in which fictions have been conceptualized, with particular reference to our three paradigms – the *PMB*-model, the *O*-model and the *S*-model. If our objective is to have an account of fiction that is mathematically mature and coherent, as well as conceptually adequate, it is necessary to begin with the issue of conceptual adequacy, which is prior. It bears remarking that to date most of the conceptual work on fiction has been produced independently of attempts to develop formal structures in the manner of a level 3 approach.

4.1 PMB approaches

PMB approaches invoke a pair of inequivalent notions. One is pretense and the other is make-believe. In some variations of the model, the two concepts are hand in glove. Even so, as we shall later see, their disjointness should not be overlooked. Most examples of the *PMB* approach arise from what we might call a *propositional attitude* orientation towards the analysis of fiction.[30] Here the basic idea is that fundamental to the creation of fictions and to their being responded to as fictions is the role of the imagination, where imagining is construed as a propositional attitude. Not all propositional attitude approaches conform to the *PMB*-model, but *PM*-approaches are all propositional attitude approaches.

By far the currently dominant *PMB*-theory is that of Kendall Walton in his widely acclaimed *Mimesis as Make-Believe* [Walton, 1990]. There is also something to be said for the stronger claim that Walton's is widely considered the best conceptual account of fiction — the best of the *PMB*-type and better than what is on offer in any of the competing models. My interest here is with Walton's theory as a paradigm of the *PMB*-model. Since Walton is to some extent anticipated by John Searle's paper, "The logical status of fictional discourse" [Searle, 1975], it is best to begin our discussion there.

4.1.1 Authorial pretense

According to Searle, pretense is the characteristic *authorial* attitude towards the sentences they pen in story-telling. When Doyle first wrote "'Good!' said Holmes. 'Excellent!'", he made no assertion. Rather he pretended to assert that this is what Holmes said. Similarly, when Doyle produced the sentence, "Holmes stretched out his hand for the manuscript and flattened it upon his knee", he pretended that there existed and that he had referred to a real Sherlock Holmes. What makes this possible is that by “pretending to refer to people and recount events about them, the author creates fictional characters and events.” [Searle, 1975, 33]. It should be noted that here Searle’s way of being a *PMB*-theorist takes on aspects of the *O*-model. For, unlike Walton, who holds that Sherlock Holmes is an object of no kind, Searle allows that Holmes is an object of some kind, namely, the fictional kind. Equally, whereas Walton holds that pretending to refer is reference of no kind, Searle’s view is that in pretending to refer we do refer, albeit to something of a kind different from the pretended referent of the pretend-reference. This gives us what we might call

---

[30] Other propositional attitude approaches include [Smith, 1978; Gabriel, 1979; Oversteegen, 1982; Currie, 1990; Mooij, 1993] and [Everett and Hofweber, 2000].

PROPOSITION 17 (Searle’s Principle). *Suppose that x doesn’t exist and yet pretending to refer to x as real succeeds. Then in pretending to refer to x, one is in fact referring to an unreal x.*

Searle himself does not explicitly proclaim *Searle’s Principle*. But he is committed to it. This being so, it is especially awkward that nowhere in [Searle, 1975] do we find a discussion of the principle’s putative truth conditions.

If *Searle’s Principle* leaves the ontological status of the fictional Holmes in the dark, also problematic is the following pair of issues.

- What is the *reader*’s role? Do readers engage in pretense? Is it the same pretense as authors are involved in? If so, what explains this remarkable concurrence? If not, how is co-reference possible?

- What is the *critic*’s role? When someone says that Jeremy Brett captured Holmes’ character better than Basil Rathbone, is the critic engaging in pretense? If so, what is he pretending? If not, then must we not concede that pretense is not a condition on reference to fictional beings.

4.1.2 Appreciator make-believe

[Walton, 1990] is a more developed propositional attitude analysis than Searle’s. While there are clear similarities between the two accounts, there is a notable difference. Searle concentrates on an author’s attitude towards the sentences he produces. Walton emphasizes the attitudes of the reader. In so doing, Walton takes himself to have adapted the model of a child’s game of make-believe. Accordingly, in reading a story, readers are *playing games*. As in various kinds of games (but not all), in playing the game of reading, readers make use of special kinds of *props*. What is more, the props of which readers avail themselves are those that are (tacitly) *authorized* by the type of game that story-reading is.

Walton sees games of make-believe as governed by rules. In the case of the games that constitute the reading of a story, the associated rules allow that certain propositions can be imagined. These are the fictional truths of the game. Walton also thinks that a prop in a game of make-believe is, as provided by the rules of the game, an object that *generates* these fictional truths. In the case of the game of fiction, the role of prop is played by the work on the text of the story. Accordingly, PROPOSITION 18 (Waltonian works of fiction). *A work of fiction is an object whose function is to serve as a prop in a game of make-believe of the requisite type.*

What is more,

> A thing may be said to have the function serving a certain purpose, regardless of the intentions of its maker, if things of that *kind* are typically or normally meant by their makers to serve that purpose [Walton, 1990, 52].

Walton allows that stories make-true sentences in the form "It is true-in-t that S", where t is a fiction and S a fictive sentence. He gives their truth conditions as follows:

PROPOSITION 19 (Waltonian truth). *A sentence of the form "It is true-in-t that S" is true iff fiction t is such that one who engages in pretense of kind K in a game authorized for it makes it fictional of herself that she speaks truly. [Walton, 1990, 400].*

An authorized pretense is a pretense of the requisite kind. But what is a pretense of the requisite kind? As Walton sees it,

PROPOSITION 20 (Requisite pretense). *A speaker displays the requisite kind of pretense when in pretending to assert a fictive or metafictive sentence he actually asserts the Waltonian paraphrase of it.*

COROLLARY 20(a) Speakers producing such sentences other than in the "spirit of pretense" fail to display a pretense of the requisite type, but only go through the motions of so doing [Walton, 1990, 402-404].[31]

Walton's account is shot-through with obscurity. The problem is so bad that it calls it sharply into question as to how *Mimesis and Make-Believe* could have attained and preserved its status as the leading conceptual account of fiction. The obscurity problem arises as follows. Central to Walton's account are the terms *pretense, make-believe* and *prop*. As used within the theory, it is arguable that not one of these terms retains its customary meaning, yet their nonstandard, theoretical, uses are unexplained. Central as they are to the Waltonian analysis, they give rise to

PROPOSITION 21 (Walton's dilemma). *Concerning the terms "pretense", "make-believe" and "prop", one of two things is true. (1) In their application to Walton's theory, they lack their ordinary meanings, and their theoretical meanings are unexplained. (2) They have ordinary meanings incompatible with the claims of the theory.*

Let us begin with *make-believe*. Make-believe is what actors do to audiences (and children, in certain of their games, do to one another). Christopher Plummer *plays* John Barrymore in ways that get the audience to experience the action on stage as Barrymore's own. Plummer *gets* his audience to *believe* that this is happening to Barrymore. Hence *make*-believe. One must not gainsay the genuine complexity of the relationship between Plummer and his audience. In getting his audience to experience his behaviour on stage as Barrymore's and to care about it as Barrymore's, something quite remarkable has occurred. We might say that something similar occurs between Doyle and his readers. Doyle's sentences get readers to experience the events described as happening to Holmes and to care about them as affecting Holmes. Yet not a word of *Mimesis and Make-Believe* suggests that this is what Walton means by make-believe.

---

[31] Are they then pretending to pretend?!

In a further of its common uses, "make-believe" carries connotations of the unreal. It ranges from Plummer's state of mind (make-believe grief), to models (make-believe markets), and from them to props (a make-believe gun). But there is nothing here that comes close to elucidating Walton's idea of fiction as a game of make-believe. Of course, there *are* games of make-believe. Cops and Robbers is a game of make-believe. But it is folly to liken the reading of a story to Cops and Robbers. In playing a game of Cops and Robbers it is not essential to its success (indeed it may be fatal to its success) that spectators be doxastically and affectively engaged — there is nothing here like being *convinced* that Hamlet is a homosexual or *hating* Iago's guts.[32] Fiction stands apart from this in two respects. One is that *readers* play no role remotely similar to that played by any of the parties to a game of Cops and Robbers. The other is that a story fails if it fails to engage its readers doxastically and affectively.

Another vexation is Walton's appropriation of the idea of a *prop*. "Prop" is a theatrical term. It is short for "properties". A prop is an object from the company's properties department. The task of the Royal Shakespeare Company's properties department is to dress the stage and equip the players with the objects and other physical wherewithal required by the play. In most cases these are not only real objects, but real cases of the type of object that the play calls for. When the play calls for a dinner table to be set, the props department sets it with real plates. Similarly, curtains are usually real curtains, chairs real chairs, beds real beds, and so on. There is in the idea of a prop not the shred of a presumption of type-unreality.

Sometimes the objects called for by the play are too big, too heavy, too expensive or too dangerous for use on stage. In that case, the props department will furnish stands-in for the objects not present. An object stands in for an object not present by virtue of various similarity relations, but also by virtue of its place in what is played out on stage. So a cap-pistol can also be a prop, standing in for a .45. Equally, the stool that is broken over the actor's head can be a prop, standing in for the real thing, which, had it been present, would have surely ended the performance. Although some props are pretend-objects, such is only a contingent and occasional connection. Most props are not pretend-objects of any kind. This makes Walton's appropriation of the term rather perplexing.

Walton thinks that *The Hound of the Baskervilles* is a prop. Like the .45, it is a real object. Like the .45, its "function" is to generate truths. In the performance of a play, the actor playing Spike might be brandishing a .45, thus helping to engender the truth that *Spike* is brandishing a .45. The text of *The Hound of the Baskervilles* also engenders a truth. In Walton's telling, it is true-in-*The Hound of the Baskervilles* that Holmes was a detective because the prop, *The Hound of the Baskervilles,* is such that one who engages in pretense of the requisite kind in a game authorized for it makes it fictional of himself that he speaks truly. Now whatever this might actually mean, it isn't a jot necessary or helpful to invoke the concept of prop. Of course, it can't be ruled out that Walton has a theoretically loaded sense of prop in mind. If so, he might have let us in on the secret.

---

[32] There is a large literature on fictionally induced emotions. See, e.g., [Radford, 1975; Slater, 1993; Walton, 1990] and [Alward, forthcoming].

Pretense is central to Searle's account of the author's role, and to Lewis' as well [1978; 1996]. Pretense is central to Walton's account of the reader's role. According to Searle and Lewis, authors are taken to be pretending to be reporting the actual truth of what they are writing about. Upon reflection however, there isn't the slightest reason for thinking that this is so. What *is* true is that when an author pens his lines, he is writing a story, and that writing a story is a *special* kind of make-believe. The author writes the story in the hope that its readers will be made to believe that what it ascribes to its characters is actually happening to them. It is, again, exceedingly difficult to fathom the mechanics of such make-believe, but there is no denying its efficacy or its power. People argue with the greatest intensity about Hamlet's character and motivation, and are reduced to tears at Little Nell's brutal death. If the *ordinary* meaning of "pretend" is to have any sway here, we should have to accept that if these make-believes arose from the author's pretense, then it would have been the case that in creating the story, the author was intent on deceiving us. If he was pretending to report the actual facts of the case, he would have been deceiving us; and, for this to be true, we would have to think that he was reporting what actually happened. It matters greatly that authorially intended make-beliefs succeed wildly even in the face of his readers' knowledge of what he has done, namely, write a story. So, for the present sense of "pretend", the last thing that the author is doing is pretending to do what he is not doing in fact.[33]

Of course, other senses of "pretend" might also be considered. When, in a game of Cops and Robbers, little Harry pretends to be Dick Tracy, he is not trying to deceive his playmates that he is indeed that pointy-chinned sleuth. He is play-acting. He is playing the role of Tracy. It is a condition of his doing this well that his playful behaviour be somehow indicative of Tracy's behaviour (he can't play him as the Virgin Mary).[34] When Sir Arthur Conan Doyle sat down to write *The Case of the Speckled Band*, he was not pretending to be reporting to his readers the adventures of Holmes. This is not to say that in this story the adventures of Holmes are not reported at all. In fact, they are reported by Watson. Somehow Doyle gets it to be the case that the events recounted by Watson are true of Holmes;[35] but, however this was done, it was not done by Doyle pretending to report those doings as fact, as Lewis expressly claims. Neither was it done by Doyle play-acting that he is reporting the actual fact that Watson is reporting the actual facts about Holmes (and about himself; these stories are Watson's memoirs).

Pretense, in both of the two senses under review, leaves no metaphysical footprints. What Doyle was doing was writing the Holmes stories. He was making them up. Doing this also leaves no metaphysical footprints. In writing his stories, Doyle constructed narratives. It is quite common to use narratives to deceive people, to lie to them. It is not uncommon to use narratives for make-believe. But it is a massive disanalogy to equate story writing with pretense in either of these two senses, just on the grounds that the narratives of fiction don't leave metaphysical footprints.

---

[33] Still, Walton is in celebrated company. Plato makes the same mistake in the *Republic*, in making the case for the exclusion of the dramatic poets from the perfect civil society. Plato was entirely right to notice in the arts a propensity to corrupt, but was entirely wrong in identifying the corruption as a form of lying.

[34] Perhaps he might play Tracy as *Tracy* pretending to be the Virgin Mary, but little Harry would have to be one sophisticated little kid to do that.

[35] This is challenged by Holmes himself in the successor novels of Laurie R. King. An interesting complication.

The idea that authors are pretenders, in either of its two senses, has nothing to recommend it. The kindred suggestion that readers also play such a role is, if anything, even less worthy of a gentle reception. Readers might misread a work of fiction, taking it as an historical narrative, although such confusions are much rarer, and harder to fall into, than one might think. But when a reader reads a work knowing it to be a work of fiction, there is a certain psychological state that it is necessary for him to be in, a certain mode of concomitant susceptibility. He must be moved to take the story's claims as true of its characters and he must be capable of being affectively engaged by those characters and the events that befall them. No doubt the full story of this psychological complexity is a long way from being told. But we know its rudiments. The appropriate response of a reader to a work of fiction is one in which he accepts the story's attributions as true recognizing all the while that the events attributed and the subjects to which they are ascribed are not real. (Hence, "the *willing* suspension of disbelief.")

Another leading *PMB*-theorist is Lewis.[36] Lewis is equivocal about fiction. He produces two accounts, whose differences are apparent in the following two truth definitions. First,
PROPOSITION 22 (Lewisean truth I). *A sentence in the form 'In fiction f, ϕ' is non-vacuously true iff some world where f is told as known fact and ϕ is true differs less from our actual world, on balance, than does any world where f is told as known fact and ϕ is not true (1983, 270].*[37]

Second,
PROPOSITION 23 (Lewisean truth II). *A sentence of the form. "In fiction f, ϕ" is non-vacuously true iff whenever w is one of the collective belief worlds of the community of origin of f, then some world where f is told as known fact and ϕ is true differs less from the world w, on balance, than does any world where f is told as known fact and ϕ is not true (p. 273).*

The idea of a *collective belief world* plays an important role in the specification of maximal accounts, and we shall have occasion to return to it later. Suffice it to say for now that collective belief worlds are populated by background beliefs. A background belief is one which more or less everyone in the author's community accepts, virtually everyone believes that everyone accepts, and so on. Lewis also requires such beliefs to be jointly consistent.

It is easy to see that for Lewis, too, pretense is essential to the analysis. A fictional story is one told as known fact, but storytelling is pretense; the story-teller pretends that he is recounting a story he knows to be true. Lewis' approach is also a heavily *consistentist* one. It may be that in f, S and that in $f, \sim S$, but it cannot be true of any f and S that in f, $S \wedge \sim S$. On the face of it, this cannot be right. In *A Sound of Thunder*, Keith both is and is not elected president in the same election.

---

[36] Critics of Lewis include [Byrne, 1993; Currie, 1990; LePoidevin, 1995; Phillips, 1999; Priest, 1997; Jacquette, 2001]. Of the favourably-minded camp, [Hanley, 2004] thinks better of Lewis' account than Lewis himself does.

[37] Note the similarity between "true in f" and "T^F".

Even so, Lewis won't have it. What is true in *A Sound of Thunder* is some subset of its fictive sentences. Lewis considers two ways in which flat-out sentential contradictions might be disabled. One is the *method of intersection*, and the other the *method of union*. Much of the criticism that Lewis' theory attracts is directed at this refusal of authorially validated in-fiction contradictions. This, too, is an issue that I'll take up later. Here it is enough to emphasize the point that for Lewis authors of fiction are pretending to tell the truth, and when, in fact, there isn't the slightest reason to think that this is what they themselves think that they are doing. If it *is* true that this is what they are doing, is it not surpassing strange that David Lewis would know it but Miguel Cervantes would not, and that Kendall Walton would know it, but Henry James wouldn't have had a clue? Of course, here too there is a possible explanation of these affronting asymmetries. There is a sense of "pretense" known to Lewis and Walton but unknown to Cervantes and James (and, we may suppose, virtually everyone else) for which it is indeed the case that what Cervantes and James were doing is pretending to tell the truth. But what sense of "pretense" *is* that?

Whatever we are to take them to mean, Walton's notions of pretense, make-believe and prop are embedded in an overarching concept of *game*. It is true, of course, that some games are games of pretending and make-believe, and that some such games involve the use of pretend-objects used as props. The concept of game has achieved a high importance in Anglo-American philosophy since the 1950s. Wittgenstein's likening of the speaking a language to the playing of a game has attracted enormous philosophical attention [Wittgenstein, 1956; Kripke, 1982]. Contractarian approaches to ethics and political theory also brim with the game theory of von Neumann and Morgenstern [1944] and Nash [1950]. Logic itself has taken on a game-theoretic character, beginning with the pioneering work of Hintikka [1974][38]. It would not be going too far to say that games are by now the objects of detailed, carefully refined and rather powerful theories. Not that the analysis of games is entirely a *fait accompli*. As I write, it is still a work in progress and a growth industry.[39] The sheer importance of the philosophical analysis of games embeds a twofold attraction. To the extent that it is considered a safe harbour, it invites the working researcher to repose his own theories there, to try to effect a kind of reduction of the issues he is working on (say, the analysis of X) to the theory of games. To the extent to which the theory of games is still a work in progress, it invites the reduction of (say) the analysis of X to the theory of games also as a contribution to the theory of games (in the manner, for example, of the Stone Representation Theorem [Stone, 1937]). There is much to commend in such assumptions. If the theory of X is indeed subsumed by the theory of games, then locating X within the theory of games is an efficient way of achieving genuine elucidations of the structure of Xhood. But if, by chance, X doesn't reduce to the theory of games, then a game-theoretic analysis of Xs is at best a conceptual distortion and at worst a profound mistake.

---

[38] In [1974] Hintikka mentions that he first presented the idea of a logic as a game in his 1964 John Locke Lectures at Oxford.

[39] In November 2005, at King's College London, the prestigious De Morgan Conference on Logic was devoted entirely to game-theoretic logic.

In Proposition 16 we flagged a negative thesis about some widely received analyses of the fictional. In this regard, Walton's contributions to this approach, as indeed are all pretense theories, is a disappointment. There is straighter way of saying this.

PROPOSITION 24 (Games and Make-Do). *The employment within the PMB-model of the notion of game shows indications of Make-Do.*

There is nothing in Walton's writings to suggest that any such reduction was ever in his mind. And certainly, it would be a considerable stretch to represent *Mimesis as Make-Believe* as a contribution to the philosophical analysis of games. Still, games *are* at the centre of Walton's approach to fiction. So it cannot be dismissed out of hand that he might have been drawn to this assimilation unawares. (After all, weren't we just now considering that Cervantes and James were *pretending* unawares?) What really counts, of course, is whether the concept of game plays any kind of load-bearing role in the logic of fiction. It takes little reflection to see that there is nothing good to be said for assimilation. In none of the ordinary meanings (*pace* Wittgenstein) of "game", is what an author is up to in writing a story a game and what a reader is up to in reading a story (knowing it to be one) a game. Some philosophers are of the view that Wittgenstein greatly overplayed the similarities between games and language-use. In this they may well be right. But it cannot seriously be doubted that in conceiving of story-telling and story-reading as games, Walton has transgressed every ordinary meaning of "game". It is possible, of course, that he intends an unordinary or technical sense of this word. If so, he would have done well to tell us about it.

A further sense of "pretense" can now be considered, at the juncture where the *PMB*-model and the S-model converge. Given the sense now in play, pretense arises from one person's proposal to *say* and another person's *acceptance* of it. Its natural idioms are "let us say that ...", "we put it that ...", "we take it that ..." and the like. Such expressions are the common currency of working mathematicians, especially when introducing concepts by nominal definition (or what [Russell, 1903] calls mathematical definition, as opposed to philosophical analysis). Such sayings, puttings and takings are stipulative. In mathematical contexts, stipulation is subject to a two-phase process. First there is the form of words and the speech acts that bring the stipulated object into view. Second, there is a judgement by the intended audience (here, it is the relevant research community) as to whether this is a usable stipulation, whether it integrates with what is presently known and facilitates the unearthing of new theorems. Hilbert appears to have thought that phase one was primary and autonomous. Russell appears to have thought that in the absence of phase-two communal acceptance, whatever had been wrought at phase one would atrophy and eventually die out. The idiom of *taking* gives us occasion to re-visit the *de dicto*/*de re* distinction. We remarked earlier that in our own discourse about the fictional the *de re* mode of expression is the more natural. If we allow that what an author's stipulations make true is that

3. $T^F \exists x(\ldots x \ldots)$,

we nevertheless find it natural to *take it that*

3. $\exists x T^F(\ldots x \ldots)$

When we turn to a more detailed discussion of the S-model in §4.3, it will be important to re-acknowledge the dual character of mathematical stipulationism — its objectual or quasi-objectural orientation and its purely instrumental orientation. Even so, the present take on taking is worth keeping in mind.

4.2 Meinongian logics

One of the basic approaches to the O-model of fiction is to treat fictional objects as a special subclass of *Meinongian objects*. Roughly speaking, a Meinongian object is anything that might be the object of someone's thought. Such objects are separate from the content of the thoughts that "intend" them. Those that inhabit space and time are said to exist. Those that have an abstract character, such as numbers or the propositional contents of sentences, are said to subsist. But there are also objects, such as Sherlock Holmes, that neither exist nor subsist. Some objects, such as Berkeley's round square are impossible. Some objects are incomplete ("... there is no fact of the matter about Inspector Lestrade's blood type." [Lewis, 1983, 270].) Some impossible objects are also incomplete, (the round square, again, which is neither blue nor not blue.) Meinongian ontologies are of considerable interest to O-minded logicians of fiction. Two of the best Meinongian logics are [Parsons, 1975, 1980] and [Jacquette, 1996, 2006].

4.2.1 Parsons' logic of fiction

Parsons invokes a central distinction among actual objects, or things that *are*, between those that are *existent* and those that are *nonexistent*. In this approach, correlated with each nonempty set of properties there is a distinct, specific object in [Parsons, 1975, 75]. Uniquely correlated with the set of properties {p: Madame Curie has p} is Madame Curie, and uniquely correlated with the set {goldenness, mountainhood} is the golden mountain. In the case of both existent and nonexistent objects, the set consists of those properties that the correlated object has. If no restrictions are placed on the properties that can be members of the sets correlated with objects, severe difficulties arise. In addition to the golden mountain, a nonexistent object, there would be a nonexistent object correlated with the set {goldenness, mountainhood, being existent}. And in addition to the round square, an impossible nonexistent object, there would be the possible round square. To avoid such difficulties, Parsons restricts the properties of sets with which objects, existent and nonexistent, are correlated to what he calls *nuclear* properties. Nuclear properties are ordinary properties of objects such as roundness and being red, whereas *extranuclear* (or nonnuclear) properties are special properties such as being existent, being possible, and being thought of by Meinong. Fictional objects — objects "native" rather than "immigrant" to fiction — are, on Parsons' view, a subclass of the class of nonexistent objects. They are the objects correlated with sets of the following sort:

> {p: according to (the 'maximum' account of) the Conan Doyle stories, Sherlock Holmes has p} [Parsons, 1975, 80].

More generally, where x is a fictional object:

> x = the object correlated with $\{p$: according to the relevant body of literature, x has $p\}$.

One final feature of Parsons' view worth pointing out concerns the relations that hold between nonexistent objects native to fiction and existent objects that are immigrants to fiction. Parsons concedes that when a fictional objects stands in relation to an existent objects, the existent object does not stand in the corresponding relation to the creature of fiction. Even if Holmes has the property of having had tea with Prime Minister Gladstone, Gladstone will lack the property of having had tea with Holmes [Woods, 1969; Parsons, 1980, 60].

Parsons' assumption that a distinct, genuine and well-individuated object is correlated with each nonempty set of nuclear properties is problematic. Some readers might find this an over-generous criterion of objecthood, and others could wonder whether, say, *unit* sets of properties are ever property-rich enough adequately to individuate. Not only is there no particular reason to think that there is just one object that is red and has no other (nuclear) properties — corresponding to the set {being red} — it is dubious that redness alone ever makes a substantive enough contribution to its bearer so as to confer genuine individuality upon it. What is more, when imported into Parsons' account of fictional objects, this assumption implies that characters who receive identical descriptions in a given work of fiction are the same character in that work [Howell, 1983, 170].[40]

A further difficulty concerns Parsons' claim that fictional objects are ontically incomplete[41]. For all properties such that neither they nor negations are in Holmes' representing set, Holmes is indeterminate with respect to those properties [Parsons, 1975, 84]. But on the face of it, it is not credible to say that Sherlock Holmes neither lacked nor possessed an alimentary canal, a liver, ten toes, two elbows or a mother; that he neither went to school nor not, that he neither did nor did not comport himself with Watson and the world in ways not expressly chronicled by Doyle. A view that rendered the indeterminacy of

> Holmes had a mole on his back

and

> Holmes lacked a mole on his back

in ways compatible with the *truth* of

> Holmes either had or lacked a mole on his back

might well be acceptable. But Parsons' account fails to meet even this weaker condition [Howell, 1979, 136].

Roughly speaking, nuclear properties are constitutive properties, such as a thing's physical attributes. Nonnuclear properties do not fix the object's kind. These include self-identity, existence, possibility and the like.[42] So, whether numbers exist or not is something that leaves it undisturbed that the kind of thing they are is numbers. The nuclear-nonnuclear distinction is important for characterizing impossible and incomplete objects. An object is said to be impossible when it has a nuclear property and its complement. An object is said to be incomplete when it lacks a nuclear property and its complement. The distinction itself is not clearly determined[43] and it is a matter of controversy as to whether some of the properties Parsons counts as nuclear really are ordinary properties of objects at all [Howell, 1983, 167-168].

---

[40] There is also the question of whether Parson's luxuriant ontology isn't occasion of unintended consequences, such as confirmation of the ontological argument for God's existence [Woods, 1986].

[41] Similarly, while it is true that Holmes lived in London, it seems not to be true of London that Holmes lived there. The idea of objectual incompleteness originates with [Meinong, 1915]. It is also a theme developed in [Ingarden, 1931] and, quite recently in [Akiba, 2000]. *Cf.* [Parsons, 1982; Lear, 1982; Simons, 1999].

Additional difficulties arise as a result of the asymmetries Parsons posits in relations that obtain between existent and nonexistent objects. Although in a fictional work the very same properties may be attributed to an immigrant object as are attributed to an object native to the work, the latter will have but the former will lack these properties [Howell, 1983, 171]. These asymmetries pose intractable problems for an adequate account of transfictive discourse — claims about the relations of characters and events that occur in works of fiction to existent things external to the works in which they occur.

In certain respects Jacquette's is an improvement on Parsons' treatment of *Meinonge* (after all, Parsons has broad shoulders). A particular advantage is that Jacquette offers a formal definition of nuclear and nonnuclear properties. He also furnishes the reader with a plausible, though informal, means of making the distinction between them recognizable. The formal definition draws on a distinction between sentence negation, represented by $\sim$ and predicate complementation represented by *neg*. If S is a sentence, its negation $\ulcorner\sim S\urcorner$ is a sentence. If Q is a predicate, its complement *neg* Q is a predicate. If Q is a nuclear-property predicate, then negation and complementation are irreducibly different. If κ is a nonnuclear predicate, then sentence negation and predicate complementation are everywhere interdefinable. [Jacquette, 2006, section 5.1]. A property is nonnuclear if any only if it is definable via logical operators and uninterpreted predicate symbols and nothing else. A property is nuclear if and only if it is not nonnuclear.

Jacquette's Meinongian logic is a quantificational extension of the basic Łukasciewicz 3-valued system. As remarked earlier it is consistent, sound and complete. Compactness, deduction and extrapolation theorems are provable. Much of classical logic is preserved. The set theory of the logic's semantics is nonstandard and paradox-free, notwithstanding that it has an unrestricted comprehension axiom.

4.2.2 *Jacquette's logic of fiction*

> Fictional objects arise from the creative endeavors of their authors. As a story is created, the author determines by free assumption the nuclear properties of its characters. These are what John Woods in *The Logic of Fiction* calls the characters' 'sayso' properties. [Jacquette, 1996, 256].

---

[42] The nuclear-nonnuclear distinction originates with Mally, not Meinong. See [Rapaport, 1978] and [Zalta, 1983, xi-xii].

[43] Another hard case: the property of being fictional. How plausible is it that being fictional or real is immaterial to whether Holmes is or is not a man?

What is more,

> The author is the primary and often the only source of what nuclear properties characters do or do not have. (p. 257)

On the face of it, this is rousing endorsement of axiom *D*, the author's sayso axiom. But appearances to the contrary, nothing remotely like axiom *D* is sustainable in a Meinongian theory of objects. Let Σ the set of all and only the nuclear properties of male human beings. Holmes, as it happens, is constituted by a particular subset of those properties. But since Doyle did not create those properties, and those properties constitute Holmes, it cannot in any strictness be said that Doyle created Holmes. Given the way that Meinongian truth works can it be the case that whereas the objects of fiction precede an author's creative intervention, the truths about those objects are precisely what are conferred by it. Every truth about Holmes is fixed by the properties he has and the relations he bears to others, and others to him. Not only do the properties constitutive of Holmes precede authorial intervention, so do the properties constitutive of Holmes' history. So a Meinongian logic of fiction violates *D*, the sayso axiom.

[Woods, 1969] draws attention to the previously remarked apparent asymmetry in relations that normally one would have taken as symmetric. If it is true of Holmes that he lived in London, why would it not also be true of London that Holmes lived there? But surely it is *not* true of London that Holmes lived there. Certainly "... a stake-out on Baker Street would have obtained no trace of Holmes" [Routley, 1981, 563]. Jacquette cites with approval Routley's own proposed solution, which he calls "contextualism". The sentence, "Holmes lived in London" is analyzed as "According to the detective stories of Arthur Conan Doyle it is true that Holmes lived in London", and the sentence "London was where Holmes lived" is analyzed as "According to the detective stories of Arthur Conan Doyle it is true that London is where Holmes lived". Routley's intuitions tell him that the first paraphrase is true and that the second paraphrase is false. Jacquette's intuitions confirm this solution. I can only say that I am at a loss to understand these intuitions.[44] Here is why.

Routley's device of contextualization takes many forms in the literature. One of the most prominent of these is the postulation of the fictive operator T^F carrying essentially the Routleyian meaning "it is fictionally the case in story *F* that" or "it holds in story *F* that". Routley's paraphrases are meant to elucidate our pre-theoretical impressions that "Holmes lived in London" is true and "London is where Holmes lived" is false. But "According to the detective stories of Arthur Conan Doyle it is true that Holmes lived in London" is no warrant that Holmes *did* live in London unless Doyle's sayso is authoritative. It would be authoritative if Doyle's sayso *made* it true that Holmes lived in London. But, as we have seen, that is the last thing that Doyle's sayso can do if the Meinongian theory of objects is correct and Holmes is a Meinongian object.

---

[44] [Woods, 1969] proposes a different solution, by invoking a distinction between what is *true* of an object and what is *true and history-constitutive* of it. Accordingly, that he lived in London is true and history-constitutive of Holmes, whereas that this is where Holmes lived is true but not history-constitutive of London. The two sentences are true together and false together but they differ with respect to history-constitutiveness. Symmetry is preserved in the first instance, but not in the second . Looking closely at Routley's example of the false member of the pair, we see that he blurs this very distinction. "According to the detective stories of Arthur Conan Doyle it is true *of* London that it was lived in by Holmes", equivocates as between "it is true that London was lived in by Holmes" and "it is history-constitutive of London that it was lived in by Holmes". The distinction is pre-figured in Frege's notion of *Sinn*. Roughly speaking the *Sinn* of '*r*' gives that part of its history that individuates *r*. Accordingly, "was lived in by Holmes" is not part of 'London's' *Sinn*.

This leaves wide open the question of how to interpret the authorial role of Doyle. Doyle may be the primary source of *information* about Holmes, but, on the Meinongian account, he cannot be the source of the *truths* about Holmes. Properly understood, the author is reporting facts that the antecedently present Holmes already has. Yet it seems equally essential that whatever else might be the case, the author not take himself to be doing this; rather that he take himself to be satisfying the sayso axiom D. It is not impossible that this is what happens. Perhaps literary authorship is indeed a kind of delusion in which, in thinking that he is making it all up, an author chances upon a set of antecedently constituted facts, which deluded readers, in their turn, take as facts constituted by the author. The delusional theory of literary creativity cannot be dismissed out of hand. For one thing, it is almost certainly the only safe harbour for Meinongians. However, perhaps one might be permitted to observe that, true or false, the delusional theory must be the most rococo theory of literary creativity on record.[45]

The nuclear-nonnuclear distinction is fundamental to the Meinongian approach to *fictional* impossibilia. What is wanted is a conception of the logically impossible whose positive instantiation does no violence to the law of noncontradiction. Accordingly, that the round square is both round and square implies that it is both round and not round and square and not square. But "x is round and not round" is formalized not as "x is round and $\sim x$ is round", but as "x is round and *neg* x is round". Since "x is round and *neg* x is round" does not imply "x is round and $\sim x$ is round", the round square may safely be constituted with no disturbance to the law. Since fictional objects are special cases of Meinongian objects, they too are subject to these same constraints. While some fictional objects are *impossibilia,* the law of noncontradiction remains intact.

As we have seen, other theories of the fictional allow for logically impossible fictional objects without the necessity to call upon a distinction between sentence-negation and predicate-complementation. In these cases, a preferred device is the fictive sentence operator T^F ("It is true in story F that … "). T^F operates somewhat like a modal operator. It allows not only chat

$$T^F(S)$$

and

1. $T^F(\sim S)$

but also that

2. $T^F(S \wedge \sim S)$

---

[45] When we were considering pretense theories in the manner of Walton and Lewis, we reflected on the possibility that in writing *Don Quixote*, Cervantes was deluding himself that he was not pretending to tell the truth. The delusionality posited by the present case is oppositely directed. What is now under the glass is whether, by Meinongian lights, Cervantes was deluding himself in thinking that he *wasn't* telling the truth.

none of which violates the law of noncontradiction. It will not have escaped notice that Meinongians such as Routley and Jacquette are also drawn to a device something like the fictive operator, expressed by the clause "According to the story [ies] of so-and-so it is true that … ". With this at hand, it is difficult to see why fictional *sentence*-contradictions cannot also be allowed, since they are expressible in the form

4. According to the story [ies] of so-and-so it is true that $S \wedge \sim S$.

On the face of it, sentences embodying the Routley modality are sentences that acknowledge the authority of axiom *D*, the author's sayso axiom. If that were actually so, then sentential contradictions are true *in fiction,* and are so without disturbing the repose of the non-ontradiction law. Certainly such would be the dominant view of the contradiction about Keith being elected and not elected President in one and the same election. But in the hands of Meinongians, Keith is an item subject to whatever constraints are imposed on all objects by Meinong's theory. One is that in *no* sense does any object violate the law of noncontradiction. So it is pre-determined that if indeed Keith is a Meinongian object, Keith was not concurrently elected and not elected President, notwithstanding that precisely this is what happened in *A Sound of Thunder*. Accordingly, the truth of "According to Bradbury's story it is true that Keith was elected and not elected President" is wholly irrelevant to what Keith did or did not do. But if that is so, the Routleyan device "according to the story [ies] of so-and-so, it is true that …" is *also* inert in the context of symmetrical relations that lose their symmetry in fictional contexts. This takes us back four-square to the sayso axiom. Not only do authors not validate contradictory goings-on among the fictional, there is nothing whatever that they make true of them. What, then, is to be said about the asymmetry between it being true of Holmes that he lived in London, yet not true of London that this is where Holmes lived? The original intuition of [Woods, 1969] comes close to being right for Meinongians. For, by Meinongian lights, Holmes is constituted by his properties, and London by its properties. Holmes is constituted in part by the relational property of living in London, but London's counterpart property is not that of having been lived in by Holmes. It is rather the property of having been *represented* as having been lived in by Holmes. There appears to be a general rule for this.

PROPOSITION 25 (Asymmetries). *For any x and y, if x is a fictional object and y a real object, then if it is constitutive of y that it is represented as bearing R to x, then it is constitutive of x that it does bear R to y.*

As was originally surmised, we now have the wherewithal to represent in a natural way the intuition that, whereas Holmes lived in London, he really didn't. True by the author's sayso are

5. Holmes lived in London.

6. London is where Holmes lived.

History-constitutive by the author's sayso is

7. Holmes lived in London.

History-constitutive by the *up-shot*[46] of the author's sayso is the sentence

7. London is represented as the place where Holmes lived.

But what is *not* history-constitutive by the author's sayso or its upshot is

8. London is where Holmes lived.

4.2.3 What authors create

The Meinongian ontological constraints have the force of restricting an author's creative wherewithal to making stories rather than making true the propositions which the story advances. So constrained, authors create neither their objects nor the truths about them. Authors create *stories* about such objects in which various such truths are *told* of them. Accordingly,

PROPOSITION 26 (Meinongian creativity). *Given that fictional objects are Meinongian objects and that what is true of them are Meinongian truths, an author's creative endeavour is wholly restricted to making fictions of such objects by the telling of such truths.*

COROLLARY 26(a) Fictional creativity has neither ontological nor alethic import. Therefore, it is necessary to read "T^F" more on the model of "according to the Associated Press" rather than on the model of "according to the Supreme Court".

With Proposition 26 in hand, Meinongians would appear to have the wherewithal to answer the objection that in their failure to satisfy axiom *D*, Meinongian theories are just wrong. For it now be rejoined, with some plausibility, that, as we presently have it, there is something quite right about *D* and yet also something wrong. What is right about it is that there is something that the author freely creates. What is wrong about it is the presumption that what it creates are objects and truths about them, when in fact all that it creates are *stories*.

This is a powerful suggestion. It has an unmistakable strategic appeal, and it generalizes in interesting ways. Not only does it offer a principled basis for rejecting the intuition embedded in axiom *D*, it also tells equally against what many regard as the foundational contrast for a logic of fiction, the contrast between non-existents and nonesuches, and the allied distinction between bet-sensitivity and its bet-insensitivity. On the present suggestion there *is* indeed a distinction between the present king of France and Sherlock Holmes. In the old way of thinking, the contrast was ontological. Nothing whatever is denoted by "the present king of France", whereas Sherlock Holmes is a somebody who happens not to exist. What is more, rooted in these ontological arrangements are explanations of additional epistemic and semantic commonplaces. We *know* who Holmes is and what he did and suffered, but concerning the present king of France, there is nothing to *be* known. Sherlock Holmes is a *well-individuated* entity, whereas with the present king of France, there is nothing that begins to count as individuating. Sherlock Holmes manages a brisk trade with the truth values, *truth* and *falsity* (both of them), whereas the present king of France is paradigmatic occasion of truth-valuelessness.

---

[46] Upshot is transfictive. The sentence, "London is represented as the place where Holmes lived" neither occurs in the text of any Holmes story nor is inferable from the sentences that do. It is inferable, in fact, from our knowledge that Doyle borrowed London for a guest-appearance in his stories.

Might there not, however, be another way of explaining these contrasts without the necessity of positing an ontic chasm between Sherlock Holmes and the present king of France? As Slater suggests (also Griffin), the chief difference between the two lies in the *contingent* fact that there are stories about the one and, so far, none about the other. They are the source of what is known (and knowable) about Holmes, they are what makes for Holmes' individuation, and they are what constitutes what is true of him. If there are explanations that we were prepared to accept, then a number of previous criticisms would lapse. In particular, it might now be possible to rehabilitate accounts of fiction in whose base logics the ontic contrast between Holmes and the rootless monarch is not honoured — whether free logics, epsilon calculi, supervaluational logics or Meinongian logics. True, the distinction between the two would have to be drawn somewhere; but presumably this would be taken care of once the story component is folded into the base logic.

Axiom *A* embodies the powerful intuition that what sets Holmes part from the present king of France is that "Holmes" refers to Holmes and "the present king of France" refers to nothing. Accordingly, there is such a thing as Holmes, but no such thing as the present king of France. If this is so, if this is indeed the difference we are questing after, it certainly *looks* like an ontic difference. There *are* things that don't exist. Holmes is one of them. The present king of France (if the *façon de parler* may be allowed) is not. If the explanations presently proposed were allowed to stand, it would appear that axiom *A* has lost its moorings. Whether this is actually so is a matter to which we shall return. Suffice it to say for now that

PROPOSITION 27 (Doubting the axioms). *As the conceptual account of the fictional evolves and deepens, the idea that the intuitive axioms are somehow sacrosanct loses some of its original luster.*

4.2.4 Worlds

[Jacquette, 1996] contains a general account of mixed sentences, in which Parsons' contrast between "native" and "immigrant" objects is put to good use. Migrancy is a considerable occurrence in fiction. Examples abound.

> We may think of Socrates in Aristophanes' *The Clouds*, Napoleon and the Battle of Borodino in Tolstoy's *War and Peace*, De Soto and La Salle and the conquest of New Spain in Edward Dahlberg's *The Sorrows of Priapus* (p. 258).

A further example is Democritus of Abdera, confined by Dante in the *Inferno* to an upper circle of hell. Migrants are objects depicted in fiction. It is the real Democritus, rather than a fictional character of the same name, who occurs in Dante's poem. Being thus depicted is one of the real Democritus' nuclear properties, as proposed in [Woods, 1969].

> Democritus has the converse intentional nuclear property of being supposed by Dante (and Dante's readers) to occupy an upper circle of hell [Jacquette, 1996, 259].

According to Jacquette, we must say that

> In creating a work of fiction, the author freely invents an alternative incomplete so-being for immigrant existent and native or immigrant nonexistent objects relative to a particular story-context or fictional world (p. 259).

Fictional worlds? A story is a fictional world? Perhaps when speaking in the casual ways of light conversation this can be said. Can it also be said with deliberate philosophical intent? No. Jacquette's logic for *Meinonge* contains a modal fragment, and rightly. If the golden mountain is golden and a mountain, then we may safely presume that it is possibly golden and possibly a mountain. If something is possibly golden, then we may safely suppose that it is not the case that it is necessarily not the case that it is golden. Jacquette adapts a standard Kripke semantics for the modal expressions of his logic. It is a semantics in which the modalities are conceived of in Abelardian fashion. Something is possibly true at a world when it is true in at least one world possible with respect to it. Something is necessarily true at a world when it is false in every world possible relative to it. At the core of this treatment is the primitive concept of possible worlds and an alternativeness relation varying through the standard relational properties of reflexivity, symmetry, transitivity and extendability. In some treatments of fiction, the careers of fictional characters play out in fictional worlds. For some theorists, fictional worlds are proper subsets of possible worlds. For others they are disjoint from them. On this latter view, what distinguishes a possible world from a fictional world is that fictional worlds can accommodate *impossibilia*, and possible worlds cannot. In some approaches, fictional worlds overlap with what modal logicians call *nonnormal* worlds. Nonnormal worlds are worlds in which everything is possibly possible and nothing is necessary (and the necessitation rule fails).

As a technical device, the notion of possible worlds has proved all but indispensable for the mathematical logic of the alethic modalities of necessity and possibility. It is well to bear in mind that giving a conceptually adequate elucidation of the idea of possible worlds is *not* a prime target of any known — or foreseeable — modal mathematical logic. Modal logicians have the same *primary* targets as classical logicians. They too, want to know what entailment is, and logical truth and consistency. Their difference is that modal logicians want to know how these properties fit languages enriched by the addition of modal expressions. It comes as no setback to these ambitions that the very idea of a possible world is the centre of a vigorous philosophical dissensus. In a recently published primer on nonclassical logics, Graham Priest touches on the several points of contention in this debate. He allows that

> The preceding considerations hardly settle the matter of the nature of possible worlds. There are many other suggested answers (most of which are some variation on one or other of the themes that I have already mentioned); and there are many objections to the suggestions I have raised, other than the

> ones I have given, as well as possible replies to the objections I have raised; philosophers can have hours of fun with possible worlds. This will do for the present though [Priest, 2001, 32].[47]

Of course it will *do*. The mathematical logics of the modals are logics of entailment and logical truth, and the like. Since these properties are now definable for modal languages, *some* attention must be given to how modal expressions might be interpreted. But the treatment of modal terms has in such logics about the same importance as the treatment of reference in the *Begriffsschrift*. In both cases they are there rather more as instrumental necessities than full-bore, conceptually adequate analyses. There is another way of saying the same thing. No one wanting to have a rich and conceptually robust analysis of reference would dream of finding it in the *Begriffsschrift*. Equally, no one wanting a rich and conceptually robust analysis of possible worlds would dream of finding it in S5 (or worse, S2).

We have been at some pains to argue that what is urgently required for a credible logic of fiction is a logic that embeds a conceptually adequate — a rich and robust philosophical analysis — of the concept of fictionality. Any attempt at achieving such an analysis by way of the philosophically half-baked mathematician's device of possible worlds is asking for trouble. It is trouble born of an over-careless fondness for devices that chance to be lying about. It suggests the menacing presence of *Make-Do*.

According to Jacquette,

> The semantic device of fictional world-indexing in modal Meinongian logic implements the story-contextual interpretation of a fictional object's nuclear properties, so that by Kripke-style stipulation Democritus [according to Dante] languishes in the inferno, not in the actual world, but in an accessible fictional Meinongian world [1966, 259].

He adds,

> The same world-indexing requirements for transworld identity of existent and nonexistent native and immigrant Meinongian worlds of the modal Meinongian semantic model [1966, 259].

Actually, I think not.

PROPOSITION 28 (Good-bye to worlds). *All going possible worlds semantics leave a conceptually adequate logic of fiction significantly underdetermined. This being so, a number of our better-regarded candidates will have to be re-considered.*[48]

COROLLARY 28(a) If Proposition 28 is right, the disinclination to invoke a worlds semantics for fiction might well be attended by a corresponding inclination to reconsider the value of conceptualizing T^F as a modal operator.

---

[47] For the present? No; for the whole book, as it turns out.

[48] Not only Meinongian approaches but those of [Plantinga, 1974; Kaplan, 1973; Gabriel, 1979] and [Lewis, 1983]. The same is true of fictional worlds treatments in the manner of [Howell, 1979; Pavel, 1986] and [Woltersdorff, 1980]. In [Wolttersdorff, 1980] fictional worlds are conjunctions of possible states of affairs and fictional characters, considered as kinds rather than individuals.

4.2.5 *The incomplete and the impossible*

It is clear from the literature that intuitions differ about the incompleteness and impossibility of fictional entities. But it may safely be said that anyone drawn to the view that fictional objects are incomplete should make a case for it, and that such a case would be at its most strong if it proceeded from an analysis of the concept of a fictional object. As even a casual review of the problem of maximal accounts makes plain, a good many of the attributes that readers ascribe to fictional characters are ascribed neither by fictive sentences nor by sentences in their deductive closures. So it is not just that either Holmes had an alimentary canal or lacked one; he *had* one (also ten toes, a mother and a brain stem). The problem begins to bite with properties that have no place in the maximal account, such as having had a tummy ache following Holmes' third-year birthday party. Of course, it might actually be the case with beings such as Holmes neither to have had nor to have been spared such unpleasantness. But what would be lost if we treated these cases in more familiar ways, in the way for example that we do with Julius Caesar? Here we haven't the slightest difficulty in representing his incompleteness as epistemic rather than ontic. Of course, we say, either Caesar had such a tummy-ache or was spared it, but which it is we'll never know. What is it about the nature of the fictional that obviates this same kind of epistemic explanation? It cannot, for one thing, be that by their very natures, fictional objects are resistant to familiar inferences or, roughly, inferences from what Lewis calls "collective belief worlds" [Lewis, 1983, 273]. This cannot be, since familiar inference is precisely what maximal accounts brim with. So Holmes has an alimentary canal, since Holmes is a human animal and that's what human animal's have.

However we resolve to handle this question, we should again note that it is *pre-settled* in the logic of Meinongian objects. If we allow that fictional objects are Meinongian, then the incompleteness of the fictional is established. But it was settled by what Meinong thought objects could and could not be, not by what he thought fictional objects had to be. In choosing a Meinongian account of objects as our base logic for fiction, we risk miring the logic of fiction at level 1. For while the logic of *Meinonge* is backed by a rich conceptual analysis of objecthood (and so makes a claim on level 3), the analysis of *Meinonge* might not suffice for the *fictional*. For this reason it is insufficient to have the ontic incompleteness of the fictional settled by what is not, after all, a theory of fictions.

Another notable constraint which a Meinongian logic also predetermines precludes even the fictional violation of the law of noncontradiction. Virtually everyone who writes about the logic of fiction is a *consistentist*. For them, it goes without saying that the law of noncontradiction is unbreachable. It is not in the slightest degree unusual that this should also be Meinong's view. Dialetheism (the doctrine that some contradictions are true as well as false) is the furthest thing from Meinong's mind. Equally, however, many who write about the logic of fiction are open to examples of the Keith kind, and ready to allow that Keith's concurrent election and non-election is a sentential contradiction that is true *in fiction*. As we have noted, one of the prime motives for introducing fictive qualifiers T^F into the analysis of fiction is to accommodate just such a possibility. So, whereas

$$\text{Keith was elected} \wedge \sim \text{Keith was elected}$$

cannot be true,

T^F (Keith was elected $\wedge \sim$ Keith was elected)

is true and was *made* true by Bradbury's story. Meinongians such as Routley and Jacquette appear to give the nod to fictive qualifiers. Routley's "According to the stories of so-and-so" seems clearly to function as T^F does. Jacquette is even more emphatic, apparently.

> As a story is created, the author determines by free assumption the nuclear properties of its characters ... The author stipulates what is supposed to be true of the people, places, and events of the narrative, which by the Meinongian [principle of the] independence of so-being *makes it true that these fictions have the nuclear properties bestowed on them by their creator.* (p. 259. Emphasis added.)

We have already noted the ambiguity of this claim. For those drawn to axiom D, the author's sayso axiom, Jacquette's words invite the interpretation that authors create the objects that their stories are about, and that authors make the sentences that are true of them true. On the contrary, as we have seen, Jacquette is a Meinongian. Holmes is constituted independently of Doyle's stories, as were whatever is true of him. Doyle's "free assumption" does not extend to any object that isn't already available — or to any proposition not already true of it. And it is this, precisely, that explains why neither Doyle nor Bradbury nor whomever else can make it true by free assumption that T^F(Keith was elected $\wedge \sim$ Keith was elected.) In fact,

PROPOSITION 29 (The concurrence of T^F-truth and truth). *In a Meinongian logic of fiction, for all sentences* S, $\ulcorner T^F(S)\urcorner$ *is true only if* S *is true.*

COROLLARY 29(a) Accordingly, in Meinongian logics of fiction the operator T^F is semantically inert, which is a further reason not to construe them as modal operators.

Here, too, we are met with the problem of which is the tail and which is the dog, and which wags which. If sentential contradictions are to be barred from fiction, it is natural to prefer that this be justified by how fiction is rather than by how Meinongian objects are. Meinong's view is that if we allow contradictions to be true in fiction, we must allow them to be true without qualification. Proposition 26 tells us quite generally that this is precisely what a Meinongian theory of fictional truth does require. So we have two options in our quest to accommodate Holmes. We can capitulate to the Meinongians. Or we can look elsewhere.

We have come to a point at which Jacquette's, Walton's and Lewis' quite different positions (see [Jacquette, 2001], acquire a degree of convergence. It is a consequence of Walton's position that only real objects can be fictional (e.g., London), whereas no object whatever is Holmes. Likewise, it is a consequence of Jacquette's position that the only truths that can be fictional truths are non-fictional truths. A further point of convergence involves Lewis. Like Jacquette, Lewis is adamant that no sentence of the form $\ulcorner T^F(S \wedge \sim S)\urcorner$ is ever true. It is easy to see what draws Jacquette to this view. It is entirely a matter of his pre-commitments. Jacquette commits the analysis of the fictional to the already standing logic of *Meinonge*. The logic of *Meinonge* is a consistentist logic; hence no sentence of the form $\ulcorner S \wedge \sim S\urcorner$ is true there. By Proposition 26, for any S

such that $\ulcorner T^F(S)\urcorner$ is true, S itself must be true. Anyone thinking that the exclusion of $\ulcorner T^F(S \wedge \sim S)\urcorner$ from a theory of fiction is just a mistake can pin the blame of Jacquette's mistake on his having insufficiently adapted the logic of *Meinonge*. In Lewis' case, it is not so clear as to where the blame should be laid. Lewis proposes to handle the problem of apparently true contradictions in fiction in one or other of two ways. By the method intersection, let F be an inconsistent fiction and R a consistent revision of it. Then $\ulcorner T^F(S)\urcorner$ is true if and only if S occurs in every R of F. Where this method employs the device of revision, the *method of union* pivots on fragmentation. Where t is an inconsistent story, C is some consistent part of it. Then $\ulcorner T^F(S)\urcorner$ is true if and only if S occurs in it at least one C of F.

A good many of Lewis' critics see the expulsion of contradictions by these methods as *ad hoc* [Currie, 1990, 69]. Others complain that it mangles our concept of the fictional to insist on these exclusions [Byrne, 1993, 26]. In his agile defence of Lewis against these charges, Richard Hanley allows that

> If it's possible for it to be fictional that a contradictory-proposition is fictional, then it *might* by Lewis' own analyses be possible for a contradictory proposition to be fictional [Hanley, 2004, 120].

Hanley considers whether this possibility might be confirmed by a thought-experiment in the manner of [Currie, 1990, 69], in which we suppose that there is a story in which someone refutes Gödel's theorem. Currie thinks this example is decisive against Lewis. Who would not agree? Not Hanley. He thinks that in imagining that there might be such a story, it is as likely that one is begging the question against Lewis in calling it a story as it is that in that imagined story a contradiction is true. What Hanley demands is an *extant* story in which contradictions are made true.

Hanley's charge of question-begging is hardly to be credited. There is no known account of it in which that is what Currie is up to. In its most general sense one party X begs the question against another party Y if and only if X adopts as a premiss in his argument against Y a proposition P in fulfillment of the following four conditions.

1. P is damaging to Y's case.
2. P is unconceded by Y.
3. P is not a consequence of what Y does concede.
4. P cannot be attributed to Y as a default.[49]

Hanley is here standing in for Lewis. He represents Lewis as holding that it is a condition on storyhood that no contradiction is true in it. For Currie simply to say, "Well here's a story with a true contradiction in it", would indeed be to beg the question. But suppose Currie's reasoning were, "There is a proposition that is in the presumptive closure of other things you have said about fiction, namely, the "story" imagined in the Currie thought-experiment is actually a story. And since it is a story in which a contradiction is true, it cannot be the case that contradictions can't be true in stories."

---

[49] For example, P is not in the collective belief world of X and Y, P is not a matter of common knowledge, P is not in the presumptive closure of propositions that Y expressly concedes.

It is true, of course, that determining what is the presumptive closure of David Lewis' concessions is not an exact science. Far easier to determine what is in the presumptive closure of Hanley's beliefs. Hanley concedes that Lewis' no-contradiction policy might well fail should there be an actual piece of fiction in which someone refutes Gödel's theorem. Hanley considers a number of candidates: *Flatland, Peter Rabbit, At Swim-Two-Birds, Sylvan's Box*, the Holmes stories themselves.[50] One could reflect on these rejections with profit. But there is no need. Hanley makes it clear that if there were an extant story in which Gödel's proof is refuted, then Lewis' constraint would be imperiled. This is extremely helpful. It tells us how to proceed against Lewis.

PROPOSITION 30 (Contra Lewis). *Find an author who will make it the case that there is an extant story that executes the story-line of Currie's thought experiment — that is, making it the case that such a story exists by* writing *it.*

Suppose Hanley (or Lewis) were now to say, "But you've begged the question". Perhaps that would be the point to move on to other things.

Incompleteness is a further point of convergence between Jacquette and Lewis. Consider the sentence, "Inspector Lestrade has blood type O". Both authors allow that it might be the case that "T^F(Lestrade has type-O-blood)" is false and that "T^F (Lestrade does not have type-O blood)" is false. But they won't allow that "T^F (Lestrade has type-O blood $\vee \sim$ Lestrade has type-O blood)" is false. Both are sufficiently classical logicians to want it left undisturbed that if $\ulcorner T^F(S \vee \sim S)\urcorner$ is always true then $\ulcorner T^F(S \wedge \sim S)\urcorner$ is always false. We have surmised that Jacquette's refusal of true contradictions in fiction flow from his pre-commitment to a Meinongian ontology. We have complained that this is an arrangement in which the tail of the receiving theory wags the dog of fiction. Perhaps this was wrong. Perhaps there is another explanation that applies to Jacquette and Lewis alike. They are determined that Lestrade's indeterminancy of blood-type not overturn to the law of excluded middle, and they are sufficiently classical about logic to want the tie between excluded middle and noncontradiction to stand. Whatever we might think of such an explanation, we now find ourselves at a point at which the convergence of the two accounts blows apart. Jacquette negotiates the problem of indeterminacy without loss of excluded middle by the device of predicate complementation. Lewis shuns the device (or anyhow does not avail himself of it) and makes do with restrictions on closure, betokened by the operator T^F. There is a marked difference between the two. That the falsity of $\ulcorner T^F(S)\urcorner$ and $\ulcorner T^F(negS)\urcorner$ does not overturn $\ulcorner T^F(S \vee \sim S)\urcorner$ is made intelligible by the semantics of *neg* and $\sim$. That the falsity of the same pair does not overturn the law is also said to lie in the providence of T^F. Perhaps it does. But it is unclear as to how.

---

[50] Also to be considered is "The Mischief of Ricardo Bosque", by Djaitch da Bloo, reprinted in [Woods, 2003] at pages 226-227.

4.3 The S-model

The earliest published version of stipulationism in the Western tradition is *Genesis*. In it the Almighty *speaks* the world into being. In the beginning was the Word. No other creation story comes so close to making it literally the case that the Creator is the author of our being. In the holy writings it is also said that humans were created in the image of God. Nearly all theologians agree that, whatever the details, the fact of mankind in God's image is constituted by the sharing of certain attributes which, in the first order of things, would be considered distinctive of the divinity. Perhaps we might think that the authorship of truths is one of those similarities. Well, this is theology, and what is wanted here is logic.

Mathematics is a good place to start. Mathematicians are dab hands at speaking things into being. This, anyhow, is a fact of such presumptive weight as to stand as a pre-analytic datum for philosophical theories of mathematics. As such, it must receive the endorsement of theory or cause must be shown as to why it should be over-ridden by it. In the century just past, perhaps the most aggressive sponsor of this view is Hilbert. Hilbert famously proposed that "if it can be proved that the attributes assigned to a concept can never lead to a contradiction by application of a finite number of logical processes, I say that the mathematical existence of the concept ... is thereby proved."[51] By the existence of a concept K of something x, Hilbert means the existence of that thing *as conceived of as falling under* K. While there is room to interpret these lines in the manner of Meinong — in which case every (consistent) descriptive term would be guaranteed a denotation — it lies closer to Hilbert's actual practice to take to heart the qualification, "the attributes assigned [by the mathematician] to the concept".[52] In that case, Hilbert's view is that when a mathematician asserts or presupposes the existence of a mathematical object, it is necessary and sufficient for the truth of that assertion or that presupposition that the purported object be provably consistent. On this non-Meinongian interpretation, it is clear that stipulation plays a central role in Hilbert's philosophy of mathematics. A mathematical object exists if a mathematician consistently says so or if what he says consistently presupposes it.

Fictionalism in the philosophy of mathematics has a larger literature than can be done justice to here. Some commentators see [Vaihinger, 1935] as providing a general philosophical background, but this is far from a universally held position. Among more or less contemporary investigators, one could mention, among several others, [Balaguer, 1996; Bunge, 1997; Field, 1980; Hodes, 1984; Korner, 1967; Resnik, 1997; Tharp, 1989; Tiles, 1998; Torretti, 1981; von Freytag-Loringhoff, 1951; Wagner, 1982; Wang, 1986] and [Thomas, 2000; 2005]. It is sometimes held by mathematical fictionalists that the way to proceed with a fictionalist theory of mathematics is to appropriate, and, if necessary, adapt, a good theory of literary fiction. This, again, is the *Literary Primacy Thesis,* called into question in Proposition 5, and to which we shall return in due course.[53]

---

[51] Quoted in [Davis, 2000, 90-91].

[52] Quoted in [Davis, 2000, 90].

[53] See, for example, [Hoffman, 1999]. See also [Resnik, 1997, 184ff]: "Positing mathematical objects involves nothing more mysterious than the ability to write novels ... ".

As we have said, truth-by-stipulation lies open to varying degrees of ontological audacity. At its most careful, stipulating that P is making it the case that P is *true in some system* Σ within which the theorist makes his stipulation. Thus P is Σ-true and correspondingly describes objects that exist-in-Σ. Hyphenation is a wholly natural way of marking the stipulator's ontological caution. It is also much the standard way of sparing the teeming pluralism of present-day logic from the inconsistencies that would flow from the unhyphenated truth of all their contending inventions.

Still, as any student of even the recent history of mathematics knows full well, it is not uncommon that a portion or entire branch of mathematics might originate with a body of hyphenated truths sanctioned by stipulation together with their consequences under agreed proof procedures, only in the fullness of time to lose their hyphens en route to an established canonicity. The suppression of hyphens is facilitated by two conditions in particular. One is the omnipresence of *the realist stance*. The other is acceptance by the research community, driven by the dynamic of *epistemic integration*.

Re-visiting our earlier example, in the immediate aftermath of the death of naive set theory, Russell sought for a consistent rehabilitation of it [Russell, 1903]. Essential to his purpose was a distinction between philosophical and mathematical analyses. Russell thought that the paradox of sets that bears his name made a philosophical analysis of sethood impossible. This was because the paradox demonstrated that the very concept of set was inconsistent. This left mathematical analysis, which is what Russell actually resorted to. In its most basic sense, mathematical analysis is nominal definition, which is a form of stipulation.

Russell, in effect, saw that stipulations produced hyphenated truths, and that communal acceptance eliminates them. When this happens, as appears to have been the case with *ZFC*, the theory in question is, or approximates to being, a received theory. A *received theory* is a set of stipulated truths, and others proved from them, whose acceptance by the appropriate disciplinary community or disciplinary culture, induces in such groups hyphenation-blindness. Typically, community-acceptance embeds a successful attempt to integrate the hypthenated truths in an established branch of knowledge. Also integral to this blindness is the sheer psychological allure of the realist stance, abetted by the conditions under which later generations are *introduced* to such truths (they are introduced to them as objective mathematical facts). It is a powerful leaning. Stipulated truths are hyphenated truths. If, as sometimes happens, these become received truths, those who see them so fail to see their hyphens; in fact, fail to see their hyphens even if they are made aware of the received theory's stipulative origins. In this some may detect grave defections from rationality and honesty. We see the psychological dominance of the realist stance.[54] Realism, we said, is an especially efficient way of paying attention. Beings like us find it as natural as breathing to experience the world as external; and even when, in particular ranges of cases, it is experienced internally, the internalities are experienced as occurrences in a more capacious and embracing externality. Can't-help-it-realism is not a metaphysical theory. It is a habit of mind. Judging from its historical circumstances, its mathematical influence is often more extreme than in fictional cases. In each case, we tend to be Barconists; that is, given truths in the form

---

[54] Hyphenation-blindness generally comes in degrees. From the stipulatively bold "true-in-S" (e.g., "true-in-intuitionistic logic"), there is a further stage, which we might call adverbilization (e.g., "intuistionistically true"), thence to adjectivism ("an intuitionistic truth"), and finally, in the absence of rival abjectivisms, we have substantiation (e.g., "true").

3. $T^F \exists x(\ldots x \ldots)$

we find it natural to reformulate them as truths in the form

4. $\exists x T^F(\ldots x \ldots)$.

In one good meaning of the term, we *take* (3) as (4). But if the actual evolution of *ZFC* is anything to go on, truths in the form

3′. $ZFC\ \exists x(\ldots x \ldots)$

are now taken as

4′. $\exists x(\ldots x \ldots)$

in which one finds no occurrence of the *ZFC*-operator or of *ZFC*-hyphenation. Twice-over this should give us pause. Even in the case of something as settled and mainstream as *ZFC*, perhaps (4Σ) is a bit much. It is more than a bit much for fiction.

There are two main ways of being a Hilbertian stipulationist. One can hold that stipulations give rise to *bona fide* objects. Or one can make the lesser claim that stipulations make *reference* possible. Roughly speaking, the first way embeds a *de re* fondness for (4). The second holds back in the *de dicto* manner of (3). Each possibility has something to be said for it. In the first instance, except for their discursive origins, stipulated objects are on a par with any other created thing, whether widget or skyscraper. If pressed to explain how speaking such objects into existence is possible, the stipulationist may plead that it lies in the nature of (certain kinds of) mathematical objects to be created thus. Of course, in so saying he sets himself for Molière's dormitive virtue riposte. But aside from that, there is something rather impressive about his position. He can attach his stipulationism to the dominant view of reference, the causal theory. If reference to a thing depends on one's being an element in a chain of events, preceding links of which have had causal contact with it [Kripke, 1980], the condition is met here. For this is a view on which the object in question has a causal origination, and whose creation is a directly witnessable event.

The second is a more circumspect alternative. It imbibes the ontological caution of something like substitutional quantification. It preserves the intuition that mathematical fictions leave no metaphysical footprints. And, in seeking to preserve the intuition that sentences ostensibly about such objects can nevertheless be true, it tries to have the very cake that it eats.[55] In effect, it shifts the Hilbertian emphasis from objects to sentences, decreeing that anything provable about them is true. As a bonus, it also attaches itself to a Fregian notion of reference, according to which the reference of a term is determined by its *Sinn*, its sense, and it does so in ways that encourage a needed refinement of the general doctrine. Let 'a' be a purported mathematical object, and let P be the set of its provable sentences. Let j be the point at which a and P make their way into working mathematics. Then at j the sense of 'a' is fixed by some subset of P. If sense determines reference, the reference of 'a' is set by P.

---

[55] Cf. Azzouni's *thin posits x*, concerning which 'there are *x*s' can be true without there being *x*s [Azzouni, 1994; 2004].

One of the complaints made by causalists against Frege's account of reference is that it leaves it inadequately unexplained as to how it can be that referring terms actually pick out objects. Frege thinks that one doesn't understand a referring term 'r' unless one understands who (or what) r is, where r is whatever, in the context at hand, 'r' uniquely denotes. If the task is to know what the reference of the term is, it is clearly insufficient to understand the ordinary meaning of it (if it has one). One perfectly well understands the meaning of "the present king of France" even though it is a referential disaster. What *Sinn* contributes is, in the context at hand, are bundles of true sentences which, if simultaneously satisfiable at all, could only be uniquely satisfied. Thus those sentences make it possible to know who or what the reference of 'r' is. Since there are no true sentences about the present king of France, there is nothing to constitute the reference of "the present king of France". So it is the lack of such truths, not the lack of an object, that makes "the present king of France" a referential wipe-out. It is the same way with "Elizabeth II". Knowing who its reference is made possible by relevant chunks of her biography — her *Sinn*. It may be directly necessary for the existence of those particular truths that the reference of "Queen Elizabeth" be a real object, but it is not *directly* necessary to knowing who the reference of "Elizabeth II" is that Elizabeth II be a real object. It is precisely this that the present, more circumspect, form of stipulation exploits. In providing a common explanation of knowing what the references of, say, "the smallest transfinite cardinal" and "Elizabeth II" are, it places the question of the criterial importance of objecthood where it belongs, in the theory of truth.

On Frege's approach, reference is likened to a function, with terms as arguments and objects as values. The theory is meant to answer two questions.

1. What does it take to understand 'r''s reference.
2. What does it take for 'r' to pick out r?

A common part of the answers to these questions is *Sinn*. Concerning 1), it is insufficient that I understand the meaning of 'r'. It is also required that I be acquainted with 'r''s *Sinn*. Concerning 2), it is insufficient that 'r' has an understandable reference (*Sinn*) in order that 'r' refer to r. I know what the reference of 'Sherlock Holmes' is and I know that no object is denoted by it. The *Sinn* of 'r' does not suffice for it to be true that r is the object 'r' denotes. *Sinn* is necessary for it to be the fact that 'r' denotes r, *if* indeed r is the object 'r'denotes. By these lights, knowing 'r''s reference is one thing, and knowing 'r''s referent is another. In Frege's treatment the distinction is overlooked. Frege wanted to preserve the intuition that 'r' lacks a reference if it lacks a referent, i.e., and object which 'r' picks out. One of the valuable byproducts of the problem of fictionality is the ease with which it makes the present distinction a useful one to have. For we now have the wherewithal to make a threefold contrast between 'the present king of France', 'Sherlock Holmes' and 'Saul A. Kripke'.

1. ‘the present king of France’ lacks *Sinn* and therefore lacks reference and therefore lacks a referent.
2. ‘Sherlock Holmes’ has *Sinn* and therefore has reference, but, given its *Sinn*, lacks a referent.
3. ‘Saul A. Kripke’ is a triple-threat. It has *Sinn*, reference and a referent.

Does this beg the question against Frege? Frege thought that reference was functional and that values had to be well-individuated objects. On a quick reading, it may seem that it is precisely this feature that our present suggestion over-runs. The appearance is misleading. In Frege’s theory, whenever ‘r’ refers, it must be true that

- $\exists! x$ (‘r’ refers to x).

Reference for Frege is quantificationally rooted, and uniquely so. The same, however, is true for fiction. It is part of the *Sinn* of ‘Sherlock Holmes’ that it refers to a character generated by a fictional text. Accordingly,

- $\exists! x T^F$ (‘r’ refers to x)

We see in this trichotomy reason to invoke the distinction between objectual and substitutional quantification. In the standard case, the quantification gives an object *to which* ‘r' refers. In the fictional case, the quantification gives an object *in which* ‘r’ refers. So in both cases, objects are essential to reference, albeit objects of a markedly different kind.

It is easy to see that substitutional quantification also extends a welcoming hand to all forms of *de re* discourse about the fictional. In

4. $\exists x T^F(\ldots x \ldots)$

we have *de re* attribution without the objects. Instantiation restores T^F to left-most position. Every truth about the values of (4)’s quantifier is a truth *de dicto*.

For good or ill, substitutional quantification gives a theory of truth without objects. It is not everyone’s cup of tea, needless to say. But is very far from being dismissible out of hand; there is a lot that counts in its favour. (See, e.g. [McKeon, 2005]). It is at least a theory with legs, and as such is streets ahead of any logic of fiction yet devised.

Earlier we mentioned the perplexing fact that beings like us are capable of believing things we know to be false to be true of the fictional and of being emotionally stirred by things that we know didn’t occur. There is a tendency among literary theorists to liken these phenomena to delusions or other forms of mental relapse. It takes little reflection to see the inadequacy of such assumptions. If in watching a performance of Christopher Plummer as King lear, I slip into the delusion that this is all actually happening I cannot be said to be following the play. If, in the grip of the excitement at Reichenbach Falls, I lose all sight of its literary origins, I cannot be said to be understanding the story. It is true that an expertly performed play or an expertly crafted story is something that we can lose ourselves in. It is a condition of this happening that we understand the work that absorbs us. It is a condition on such understandings that in having them the works retain their recognizable fictionality. Interacting with a work of fiction is doing two things at once. It is believing things that one knows to be untrue and it is being moved by things that have not happened. So we speak of the things that we believe as holding in the story and of the things that move us as occurring there. Let us call such beliefs and such affects *double-aspected.* Accordingly,

PROPOSITION 31 (Essential double-aspectness). *It is a condition on one's understanding a work of fiction that the beliefs and emotions to which it gives rise are double-aspected.*

As it happens, it is an advantage rather than a liability, that a fictional name's *Sinn* should embody some sign of fictional origination. What this tells us in simple English is that you cannot know who Holmes is unless you know that he is fictional. Since this is tantamount to saying that you cannot know who Holmes is without knowing that he isn't anything, then

PROPOSITION 32 (The double-aspectness of *Sinn*). *Our understanding of the* Sinn *of fictional names is likewise double-aspected.*

Something like the distinction between the objectual and the referential senses of stipulationism attends the distinction between idealism and instrumentalism in the foundations of mathematics. The first term of the latter contrast is in play in the early debate over the philosophical legitimacy of the axiom of choice.

> On the one hand, metaphysical elements tinge the famous debate between Hadamard and French School of Baire, Borel and Lebesque — do sets exist objectively, independently of us, or do they exist only insofar as they have been defined or constructed by us? — in other words, there was a conflict between something like Robust Realism and a version of Idealism [Maddy, 2005, 355-356].

On the idealist approach, the objects of mathematics — sets, for example — are ideal objects, the fruit of the mathematicians' definitions. In one variation, sometimes called "conceptual realism", these are objects in the theorist's mind. It is possible in principle to be one or other of two kinds of mathematical idealist, or conceptual realist (shades of Kant). One might be an across-the-board idealist, holding that all mathematical objects are human intellectual artifacts. Or one might also take a realist position with regard to parts of mathematics and reserve one's idealist leanings for the more *recherché* bits. For instrumentalists, something like this latter distinction is mandatory. Some part P of mathematics is taken as privileged in some way, and some further part I — the instrumental part — is justified solely by the contributions it makes to proofs in P. One way of instantiating the $P-I$ divide is by taking P as finitistic mathematics in the sense of [Tait, 1981]. Thus finitism is number theory formalized in primitive recursive arithmetic *PRA*.[56] It is possible to proceed to I from P in two ways, instrumentally or idealistically.

> [W]e can add stronger arithmetical principles, which seems to amount to adding more induction, or we can first expand the language with variable for

---

[56] Either as an *equational* rendition or a *first order* theory. See here [Caldon and Ignjatovi, 2005].

> new types of objects (e.g., sets) and then add new arithmetical principles for these objects [Caldon and Ignjatović, 2005, 779].

The second way is the way of idealism. It also flags an ambiguity in the word "instrumentalism." The second way is also the way of instrumentalism to the extent to which the contributions of its postulated objects to its *I*-rules is justified by the *I*-rules yield in *PRA*, the privileged part of mathematics. The first way is also the way of instrumentalism, but in a purer sense. Proof rules are mere instruments. If the contributions they make to *PRA* justify their use in *I*, then, since the rules introduce no new objects, their use is entirely as proof instruments. So, roughly speaking, the difference between an idealist instrumentalism and a pure instrumentalism in extending *PRA* to *I* is the difference between instrumentally successful rules that supplement the objects of *PRA* and instrumentally successful rules involving no objects not in *PRA*.

It is well to keep in mind that both idealistically instrumental and nominalistically instrumental sets of proof instruments are judged solely on the basis of their contributions to *PRA*. While it is possible to assimilate the stipulations of the S-model of fiction to either strand of this instrumentalism, there is nevertheless an over-arching dissimilarity that might tend to count against the S-model generically. In both forms, idealist or pure, the stipulations of *I* are justified by the extent to which they facilitate the demonstration of new facts about *PRA*. For this analogy to hold in the case of literary fictions, the stipulations of *The Hound of the Baskervilles* would be legitimized only by the extent that they facilitated the production of new facts about the world. Now it is quite true that the world is made different by the fact that Arthur Conan Doyle penned *The Hound of the Baskervilles*, as well as by the fact that it has had legions of readers and critics. But, in the general case, the truths that hold in *The Hound of the Baskervilles* don't hold in the world, and those that do hold in the world were not made so to do by *The Hound of the Baskervilles*. So a question for the S-theorist of fiction is wherein does the "justification" of the truths of *The Hound of the Baskervilles* lie?

Perhaps the dissimilarity might be discounted through judicious generalization. *I* is justified by the contribution it makes. What matters most is that it is a contribution, not where it is made. No one in his right mind thinks that the truths of *The Hound of the Baskervilles* make a contribution to our knowledge of the world (to the history of Dartmoor, for example). But they do make a story. *That* is their primary contribution; that is what they are *for*. As with mathematics, there are two ways in which one might be an instrumentalist about this. One might opt for idealism and ascribe to the author's stipulations not only the truths of *The Hound of the Baskervilles* but also some new objects to serve as their subjects. Alternatively, one might elect the course of pure or nominalistic instrumentalism, plumping for the idea that whereas *The Hound of the Baskervilles* is constituted by its author's stipulated truths, these are truths that — in the manner of substitutional quantification — require no objects other than names. Here again we see some convergence. In Waltonian versions of the *PMB*-model, there are no fictional objects. The same is true of purely instrumental versions of the S-model. In these variations, fiction leaves no metaphysical footprints. With regard to a story's characters, there is nothing there.

Unless the present discussion has been seriously wrong-headed, it would appear that we have ample grounds for proposing that

PROPOSITION 33 (Giving the nod to the S-model). *Of all the going approaches, there is a version of stipulationism that is best for literary fictions.*

PROPOSITION 34 (Quantification as substitutional). *In the variation that works for fiction, the S-model allows for* de re *quantification into fictional contexts, provided that quantification is substitutional.*

PROPOSITION 35 (Reference without referents). *In this same variation, the S-model is (as it were) tailor-madet for the distinction between a term's having a reference and its having a referent, and for the allied notion that a term's reference is set by its* Sinn*, and its referent, too, but only if it has one.*

5 MAXIMAL ACCOUNTS

Maximal accounts when adjoined to a text's fictive sentences give the full story that it tells. In most of the work done to date, conditions on maximality are usually thought of as given recursively by closure clauses on sets of fictive sentences. We are now in a position to see that this is a duly restrictive way in which to construe maximality, much in the way that the lingering preference for deductive closure is an unduly restrictive way to construe fictive closure. It goes without saying that maximality owes a good deal to what follows from what in fictional settings. It owes at least as much to background assumptions. Sets of background assumptions are what Lewis calls collective belief worlds. Its beliefs are those that everyone in the community shares, more or less everyone believes everyone shares, and so on. A community's background assumptions encompass what is "common knowledge" there, including its accumulated store of lore about what things are actually like. It includes the belief that Saul A. Kripke has an alimentary canal, because it is known that Saul A. Kripke is a human and that having alimentary canals is part of what humans are like. The beliefs in a community's background assumptions come in varying degrees of strength, but over all they reflect an interesting epistemic division. By and large the generalizations are held with greater confidence than its instantiations. This is as it should be, epistemically speaking.

These observations are designed to capture the following kind of situation. We know very well that it is generally (though not universally) true that Americans are not flat-earthers. We know every bit as well that Saul A. Kripke is an American. It does not follow that Kripke is not a flat-earther, but it is something that we may safely presume on sufferance. In other words, it is a *default* that Saul A. Kripke is not a flat-earther, whereas it is a *fact* that he is a human and a *fact*, albeit a *generic* one, that Americans aren't flat-earthers.[57] So we may say that "Saul A. Kripke is not a flat-earther" is in the *presumptive* closure of a beliefs shared by readers of this chapter, and many others besides.

---

[57] Concerning genericity, [Carlson and Pelletier, 1995] is required reading. See also [Gabbay and Woods, 2005, chapter 7]

A further feature of background assumptions is that they are largely tacit. No one thinks that it is realistic for anyone whose background is that of the readers of this chapter to set out to produce a complete inventory of it. At a certain level, therefore, it is true to say that it is epistemically indeterminate as to what the contents of one's background are. Hence an ambiguity in "know" presents itself. All sorts of things that I know (by common knowledge) I cannot enumerate or specify on demand, or perhaps ever. For those cases, I do not know what I know by common knowledge. To a degree, the tacitness of common knowledge is off-set by the efficacy of retrieval-to-surface mechanisms in beings like us. A cognitive economy of tacit knowledge is like a manufacturing or retail economy run on the principle of just-in-time. But even here elements of taciticity persist. In forming the belief that, Saul A. Kripke is presumably no flat-earther, the presence of the genericization that Americans aren't flat-earthers may itself be tacit, and usually is.

Background assumptions are an attempt to carve out a portion of a usually much larger set of background *facts*. Roughly speaking, the background facts of anything I chance to know about the world is how the world actually is, whether I know it or not. Common knowledge, therefore, is knowledge of a proper subset of background facts. As measured by our own common experience, both sets are vast — well-beyond our capacity for specific direct-command manipulation. In our common experience, we haven't the slightest difficulty in acknowledging their existence and their salience. In particular, we have little inclination to ontologize their epistemic indeterminacies, or to confuse, as mediaeval philosophers would say, *ordo cognescendi* with *ordo essendi*. Accordingly,

PROPOSITION 36 (The parasitism of fiction). *Except where the contrary is indicated by a story's fictive sentences, authors tacitly include in their maximal accounts (a) the background facts, e.g., facts as to how the world actually is, and (b) background beliefs, i.e., common knowledge shared by author and reader. What this shows is the extraordinary degree to which fiction is parasitic on the world.*

COROLLARY 36(a) In making their respective passages from the world to the story, background sentences become true-in-the story, hence also T^F-true in the world.

COROLLARY 36(b) Just as the epistemic indeterminacies of backgrounds are not ontologized (in the world), neither need they be in fiction. Thus fictions incorporate complete worlds, and incomplete, though large, epistemic subworlds of them.

Maximal accounts arise from recursive interplay between and among fictive sentences, background facts, background beliefs, and consequences thereof. Maximal accounts subsume the closures of such sentences. We have already laid the stage for recognition of a four-fold ambiguity in the notion of the closures in fictional texts. Corresponding to the distinction between the consequences a story has and the consequences that should be drawn from it is a distinction between a story's *propositional* closure and its *inferential* closure. Within this contrast are the various grades of closure, corresponding to the various grades of consequence: deductive, inductive, statistical, abductive, presumptive, plausibilistic and so on. Let $L_1, \ldots\ L_n$ be logics for these consequence relations. The question is whether to embed them in the logic of fiction. There is no strategic advantage in not doing so. Let the sentences of the maximal account imply whatever they will. It suffices that they not be given a free-ride into it. In other words maximality should not be closed under consequence. This is tantamount to saying that only inferential closures should be seen as involved in the construction of maximal accounts. This is entirely the right way of seeing things, at least as an approximation. [58]

Even so, a second distinction now presses for attention. *Whose* inferences belong in the inferential closures of a story? As with the requisite collective belief worlds of a given text, Lewis thinks that it is those shared by author and reader alike. There are two things wrong with this. One is that it leaves it undetermined as to how conflicting inferences might be decided. In *Faust* it is true that Faust makes a pact with the devil. It may also have been the case that Goethe believed that the world was such that human beings could make pacts with the devil, literally. As such, this goes into the maximal account of Goethe's poem as a background consideration — as part of how the world is. Yet for many readers of the poem, this is at best a dramatically interesting superstition, never mind that they would concede that just this actually occurred in the poem. There is an answer to this. Goethe doesn't get to decide theological questions *except fictively*, just as Cervantes doesn't get to decide astronomical questions *except fictively*. Whatever Cervantes or anyone else may have thought, the motion of the heavenly bodies of part one of *Don Quixote* are elliptical, not circular, unless otherwise provided by the sentences of the text. This notwithstanding that part one was published prior to Kepler's breakthrough and only part two afterwards.

There are two questions about inferential closure, not one. Of course, we want to know what inferences we as readers and critics should draw from a text's fictive sentences and its background considerations. But there is also the question of what inferences Holmes and his ilk draw about the events of the story. In some respects, Holmes is better placed than we to fill in the maximal account. He knows more about late-Victorian London that we do, and about the myriad facts of his own life. This being so, Holmes will be privy to consequences of such facts that can only escape us. On the other hand, the fictive sentences of stories are often arranged so as to give the epistemic advantage to readers. Readers of *Oedipus* can see the trouble coming from miles away, but Oedipus has no clue until the madly bitter end. Intuitively, we want to count inferences that Oedipus doesn't make or that contradict those that he made mistakenly. To this end the nod goes to the reader. But in light of the first difficulty, it can only be a qualified nod. The best compromise would be that the requisite inferences are those drawn by anyone combining the advantages of readers and of characters within the fiction. A good, though not perfect approximation to this is the third-person narrator of stories in which reader and author are near-contemporaries.This can't be all there is to it, needless to say. Some narrators, often in first-person cases, are mistaken in various ways. In very old stories, the Homeric poems for example, there is much that I will miss that (even) Achilles will not. These are difficulties of a sort that commend themselves to the device of the ideal observer. An ideal observer is someone in the world who has mastery of a text's fictive sentences and who is counterfactually present in the story and positioned there in ways that facilitate his drawing the appropriate inferences. Because both reader and fictional character are differentially disadvantaged, readers and characters will draw different subsets of the inferences that the ideal observer is able to draw. But, for all that, these will in the general case be a substantial overlap.

---

[58] Of course, don't the facts of the world have their own consequences independently of whether we have the resources or occasion to draw them? Yes, but we may take it that the facts of the world include the facts of their closures.

5.1 Contra ideal observers

Some readers will take the resort to ideal observers as a retrograde step, if not a *cri du coeur*. Fiction is a hard enough problem (they will say) without the added burden of counterfactual in-fiction observers. For anyone still drawn to a worlds approach to fiction, the present strategy presents us with two levels of modality, one made necessary by the fictional and the other made necessary by the counterfactual, the upshot of which is that we are now required to chart the connections between non-fictional possible worlds and fictional worlds. I am myself rather sympathetic to this complaint. I have tried to say why the worlds approach to fiction is misbegotten. I admit to having similar doubts about the conceptual (as opposed to mathematical) adequacy of a worlds approach to counterfactuality. Running a possible worlds account for counterfactuals in fiction involves running a fictional worlds account of the possible worlds account that takes care of counterfactuality. This is not to say that such interactions are beyond us; but, as Quine would say, the returns had better be good. Upon reflection, it is better to be realistic. There is a great deal that is true in fiction that we shall never know. There is a great deal in the world that we shall never know. In particular, there are lots of inferences that Holmes and his ilk do (and should) draw from whatever it is that they have knowledge of. Since in lots of cases what they have knowledge of extends to what we lack knowledge of, we will not know how to infer what inferences it was right for them to draw from such knowledge. This is nothing to be over-concerned about. It is just another patch of epistemic indeterminacy in our understanding of Holmes and his doings. Nor need this indeterminacy leave a gaping hole in our appreciation of the kind of man Holmes is and the nature of his doings. For in all things, Holmes is as much like us in his nature and his actions and passions as is compatible with the fictive sentences of the stories and his status as a fictional being.

We have already seen that a stipulationist approach allows for, but does not demand, a substitutional account of quantification. Here, too, it is essential to distinguish between the quantifications we ourselves make and those made by Holmes and his ilk. Bearing on this is the parasitism of fiction on the world. To the extent that its fictive sentences allow it, quantification there should be like quantification here. There is no particular reason to attribute to Holmes substitutional quantifications except for those cases in which Holmes is trying to quantify over what he has trouble in recognizing as real. Like us, Holmes too is subject to nominalistic provocation. When this happens, we should not deny him the utility of substitutional quantification.

Holmes is a nominalistic provocation for us. Although we know who he is and what he did, we also know that he is not there and that none of what has ever happened happened to him. This is an interesting duality, for which the truth of "There are lots of things that don't exist" is a godsend, and which (on my telling) "Holmes does not exist" instantiates and "The present king of France does not exist" does not. Retaining this intuition requires a view of quantification that restricts the values of bound variables to referring terms that have sense. Perhaps this would not be a well-motivated restriction generally, but it works well here. For here truth is primary and objects are nothing.

5.2 Contra a fictive modality

In §4.2.5, we noted that in Meinongian contexts, it is hard to see what the semantic import of T^F might be. We went so far as to suggest that in such contexts T^F is "semantically inert", thereby leaving the impression that this was something that told against the Meinongian analysis of fiction. As we now see, although this may be a consequence that Meinongians would be prepared to accept, for the rest of us it is a disaster. T^F is needed to mark the fact that for any sentence in the closure of any set of sentences containing a fictive sentence is false, and yet true in the story. One of the attractions of this is compatability with the T-schema. If the T-schema is adequate for sentences about the world, it is also adequate for the sentences of fiction, except where contra-indicated by its fictive sentences. Accordingly, we have it that

> "Holmes is Watson's friend" is true iff Holmes is Watson's friend.

In other words, "Holmes is Watson's friend" is false. What is not false is that "Holmes is Watson's friend" is true-in-the story. "Holmes is Watson's friend" is true-in-the story iff it is a member of the story's maximal account M. "But that's not truth", some will complain. The reply to make is: *Of course* it is not truth, and this is nothing to complain of. If it is true that $T^F(S)$, then $T^F(S)$. But we do not have it that if $T^F(S)$ then S.

T^F is a way of keeping our assertions contradiction-free. More generally, it is a way of keeping our assertions falsity-free. In uttering $\ulcorner T^F(S)(\wedge \sim S)\urcorner$ I leave the law of noncontradiction untrifled with. This gives us no occasion to apply emergency measures from dialetheic logic. $\ulcorner T^F(S \wedge {\sim} S)\urcorner$ is true if $\ulcorner S \wedge {\sim} S\urcorner \varepsilon M$.

There remains the question of the place of $\ulcorner S \wedge \sim S\urcorner$ in fiction, concerning which a certain caution is wise counsel indeed. In "The Mischief of Ricardo Bosque", its author da Bloo provides that every sentence is true. A good part of the story involves an attempt to depict Ricardo's day-to-day reaction to the storied fact of absolute inconsistency. "The Mischief" is not, I think, a very satisfactory story on literary grounds, but it gives no offence on the score of what it obliges its readers to count as true in it. Since, every sentence S is true in the story, we must allow that for every sentence S, $T^F(S)$, where F is "The Mischief of Ricardo Bosque". However, there is nothing to indicate that Bradbury's story is one in which everything is true. If this is right, there are two options to consider.

Option (1). Allow the propositional closure of *A Sound of Thunder* to be as classical as you please; in which case, we concede that everything whatever follows from "Keith was elected President in 2055 $\wedge$ $\sim$ (Keith was elected President in 2055)". Restrict the inferential closure (and accordingly, the maximal account) dialetheically, in the manner, say, of Priest's *LP* [Priest, 1979].

Option (2). Subject the propositional and the inferential closures to the same dialetheic constraint.

It is interesting to note that neither from the point of view of what readers must concede to or from the point of view of what a story's characters must attest to is there the slightest difference between option (1) and option (2). As far as we are concerned, all that is true in *A Sound of Thunder* are the sentences in its M. From the point of view of Keith and his ilk all that is true is in the inferential closure of what they know of themselves and their situations. In more generalized contexts, there might be differential advantages in keeping one's options open, or in plumping for one or the other. But this is not something that concerns us here.

All this has a bearing on the advisability of construing T^F as a sentential modal operator. No doubt one of the attractions of so doing is that there exist well-understood semantic theories for the modals. In §4.2.4 I attempted to show that possible worlds' semantics leave the logic of fiction in a conceptually underdetermined shape. Far and away, the best to be said of a possible worlds (or a fictional worlds) approach to fiction is that we have a fairly good command of worlds when considered as abstract mathematical structures. One senses here the presence of *Make Do*. Dropping the pretense that a semantics of worlds offers conceptual elucidations of the semantic dimension of fiction, deprives the T^F-as modal-operator assumption of much of its attractiveness. For if a modal operator requires a possible worlds semantics, and a possible worlds semantics tells us nothing (conceptually) about fiction, what can a modal operator tell us about fiction?

We would be better advised to conjure up a role for T^F that more closely comports with actual practice. We should aim for a construal that allows for the propositional identity of the S that the story makes true, and the S that the reader accedes to. But it is also necessary to pay close attention to variations in truth-attribution. So while my utterance of "Keith was elected President in 2055 $\wedge \sim$ (Keith was elected President in 2055)" has the same propositional content that it has in the maximal account of Bradbury's story, it is essential that what is true there not be confused with what is true here. These objectives do not require the postulation of modalities. *Labels* will do.

Although labels have only recently been studied in a systematic way,[59] they are a far from unfamiliar device. Among logicians, labels are most commonly used in recording the derivation histories of lines in proofs, and in marking the roles played by formulas not introduced by derivation from prior lines, as in axiomatic, conditional, indirect and reductio proofs. Then, too, we saw in our discussion of the S-model of §4 that truth by stipulation gives rise to hyphenation. If a theorist stipulates the truth of S in his theory T, then all else being well, S is true-in-T. Another abundant source is conjecture. Every time a theorist completes an abductive inference, he does two things. For some proposition H, he conjectures that H is true (knowing that he does not know that it is); and, secondly, he releases H for inferential duty in the domain of enquiry within which the conjecture has arisen. In each of these cases — the theoretical stipulation case and the scientific conjecture case — the proposition in question may in time find itself wholly integrated into some or other body of *knowledge*. When that happens, H's ancestry is overlooked. But until it does happen, it is essential that its ancestry *not* be overlooked. When Planck made the momentous conjecture of quanta in 1900, he released the quantum hypothesis for further work in the physics of light. It proved a transformation of physics. Large classes of sentences were now derivable with the aid of the quantum hypothesis, and each of them retained the conjectural status of the hypothesis that abetted them until such time as independent experimental corroboration became available. Prior to that, it had to be kept in mind that all of this new physics was itself conjectural. Hence the importance of labels. Labels are an idealization of the more or less settled informal practice of not taking the experimentally unconfirmed propositions of fledgling quantum physics as demonstrably true in the world, rather than demonstrably true-in-quantum theory. Quantum theory is pretty wild conceptually, but it is not fiction. There were good reasons to want to integrate quanta into established science. When it happened, science was considerably enriched by it. With the creations of fiction, on the other hand there is nothing for them to integrate into (except, in limited cases, each other.).

---

[59] [Gabbay, 1996]. For a brief informal survey, see [Gabbay and Woods, 2003, chapter 13].

We saw T^F as a sentence operator that converts a falsehood to a modal truth, and for this to be true, we took it as given that a modal semantics would have to be dreamed up for T^F. Of course, I now see that as a mistake. T^F is not a semantically loaded modality. It is a label. This is as it should be. There is nothing semantic about the fact that whereas S is false, $T^F(S)$. S is false precisely because it is not the case that S. That is a semantic fact about $S.T^F(S)$ precisely because $S \varepsilon M$. That is not a semantic fact about S. Again, if $T^F(S)$, it does not follow, and in large classes of cases it is false, that S.

5.3 *An advantage of mathematical stipulationism?*

In [Woods and Alward, 2002] much ado was made about a problem that is fundamental to fiction and only contingently an issue for mathematics. This, we suggested, was an asymmetry of sufficient import to tip the scales in favour of mathematical fictions rather than literary fictions as offering up the more generic or canonical model, thus ending the career of *The Literary Primacy Thesis*. The putative asymmetry arises from the fact that by and large literary fictions are rivals of the world, that what they make true the world has already made false. It is true that on occasion a mathematician's stipulations will contradict some antecedently proferred theorem or definition, but by and large mathematical stipulations go where others have not yet gone.

There is a further reason to question the importance we attempted to give the asymmetry between literary and mathematical fictions. For consider all those creative stipulations in mathematics for which there was no antecedently available claim that it contradicts. Mathematics is the natural home of negation-as-failure. Any S yet to find a home in mathematics is such that $\ulcorner\sim S\urcorner$ holds by default. Accordingly, any fresh stipulation will contradict an existing default. No one thinks that a default should be impervious to contradicting demonstrations. But removal by stipulation is another matter. So we have it that the asymmetry all but vanishes. For many sentences of fiction, there is an antecedently existing fact to the contrary. For most sentences of stipulative mathematics there is a presumptively antecedent fact to the contrary. This presents us with something quite close to a level playing field when it comes to handling those contradictions. It is not a problem. We handle them with labels.

5.4 The transfictive

There is as yet no place in this analysis for transfictive truths, such as that Holmes was admired by Agatha Christie. This is a sentence that could not have been true in the absence of truths in the maximal accounts of the Holmes stories. Like certain of the sentences in the maximal accounts, transfictive sentences depend on the fictional, but with a difference. These latter are true in fiction but not in the world, whereas transfictive truths are true in the world but not in fiction. So the dependency is different.

The paradigm of a transfictive truth is

2. Reader R knows who Holmes is and what he did.

All else is variation — Christie's admiration, King's appropriation, and Brett's dramatic representation. One of the attractions of the S-model is that it suffices to know who Holmes is to know the reference of "Holmes". This is done via *Sinn*, the individuating sentences of the maximal account. A reader's knowledge of who Holmes is is a matter of his knowing with respect to those sentences S that $T^F(S)$ is true. He knows that $T^F(S)$ is true if he knows that $S \varepsilon M$. Similarly, someone knows what Holmes did when he knows who Holmes is and knows with respect to those sentences S that give the action of the story that $T^F(S)$ is likewise true. Here, too, one knows that $T^F(S)$ is true if one knows that $S \varepsilon M$.

Knowing who Holmes is different from knowing Holmes. The same is true of Saul A. Kripke. Knowing Kripke demands acquaintance with an object. Knowing who he is requires acquaintance with 'Saul A. Kripke"s *Sinn*. With 'Holmes' this is all we get. There is no object for us to acquire acquaintance of. Knowing Holmes seems to require it to be true that

$$\exists x \text{ (John knows } x).$$

But this is false. It would also seem that admiring Holmes is like this too, requiring that it be true that

$$\exists x \text{ (Christie admires } x)$$

But this too is false. How, then, can it be that Christie's admiring Holmes is modeled on her knowing who Holmes is. The answer is that it suffices for Christie to admire Holmes that she finds Holmes to be admirable; and it suffices for this that she thinks that T^F(Holmes is admirable). It is much the same way with King's appropriation of Holmes in her own *A Monstrous Regiment of Women*, and of Brett's capturing nuances of Holmes' personality. In the one case, the Holmes of Doyle's telling is the Holmes of King's. In the Doyle stories "There exists an x such that x is denoted by 'Holmes"' is true. In the King stories, "There exists an y such that y is denoted by 'Holmes"' is also true. In King's story but not Doyle's $y = x$. So some of the sentences true in King's story are not so in Doyle's.

Brett's fidelity to Holmes' personality is similarly dealt with. Two conditions need to be met. One is that Brett must know what Holmes is like. The other is that he must be able to capture these features in the way that he plays Holmes on television. Knowing what Holmes is like is knowing the requisite propositions in the form $\ulcorner T^F(K\ (\text{Holmes}))\urcorner$ And knowing how to capture K-ness is simply a matter of the actor's trade-craft.

Perhaps it will not have escaped notice that each of our intuitive axioms makes a transfictive claim. If it cannot accommodate the intuitive axioms, that will be reason to doubt the our present account. We said earlier that we should prepare ourselves for the eventuality that our intuitions will prove to have been misconceived. Certainly such was the lesson of the inconsistency that brought down intuitive set theory. It is quite true that these might have to be given up, but they should not be given up cavalierly.

Axiom A tells us that we refer to fictional beings even though they do not exist. We have already seen that the S-model accommodates a view of reference according to which one can know the reference of "Sherlock Holmes" without its having a referent. We have also seen that in combination with substitutional quantification we may give such reference *de re* formulation in the manner of $\ulcorner \exists x T^F$ ('r' denotes x)$\urcorner$.

Axiom B tells us that there are true sentences about the fictional. Again, we have seen that the S-model allows us to say that the only truth in a story are those that are true independently of it. All else is untrue. Yet everything in a story's maximal account is true-in-the story.T^Fpreserves all that is desirable in B if we are also going to allow that the truths of stories leave no metaphysical footprints.

Axiom C is about inference. I have already said my piece about inference in the earlier discussion of maximal accounts. The rough rule of thumb is that inference within fiction be as near as like inference in the world as the fictive sentences permit. In the particular case of $\ulcorner S \wedge \sim S \urcorner$ occurring in a story, the surrounding sentences of the maximal account will serve as context. In a story such as "The Mischief", one infers that everything is true in the story. In a story such as *A Sound of Thunder*, inference is paraconsistent — indeed dialetheic. A further tip: anything inferred from a sentence, labeled 'T^F' is itself labeled 'T^F'. Storied truths don't achieve cheap or unfettered *entré* to the world.

Axiom D provides that a story's fictive truths and those in their inferential closures are made true by the author. The truth is that they are made true in the story by the author. In some versions, D also has it that the author provides an ontology for his stories, a realm of nonexistent objects and happenings. There is a version of the S-model that allows for this; but not every variation does. In its purely instrumental form, all we can say (or would want to) is that authors provide ontologies in their stories — objects and events that are real in the stories, our acknowledgement of which is protected by the label 'T^F'.

Finally, E reminds us that real entities may be borrowed for guest appearances in fiction. This is true, but it is a truth of an appropriately distinctive form. For example,

$$\exists x\ (x = \text{London} \wedge T^F(\ldots(\text{London})).$$

Accordingly, whereas it is possible that

$$\exists x(x = \ \text{London} \ \wedge \ T^F(\ldots(\text{London}))$$

is part of the *Sinn* of 'London', nothing in the form of 'T^F (... (London))' is part of it.

One of the virtues of the S-model is that it preserves the flavour of the intuitive axioms while at the same time allowing for refinements driven by the foundational fact that the fictional leaves no metaphysical footprints. A further virtue is that it leaves more or less standard approaches to reference, truth and inference intact, taking care of needed qualifications by way of labels. What this helps us see is how close to a logic of non-fictions a logic of fiction turns out to be. This is as it should be. Fiction is parasitical on the world, with which it achieves a kind of partnership. But it is an arrangement in which it is very much the case that the world is the senior partner.

When we say that fiction leaves no metaphysical footprints, we nominalize the fictional, that is, we drift towards a nominalistic account of it. One of the difficulties thrown up by nominalism is that of integrating its nominalities with the tough customers of the real world. Such integration comprehends drawings of the real by the merely nominal, which is ludicrous on its face. Attempts to minimize this difficulty which trade-off ontic integration for predictive yield gets a kind of instrumentalism up and running. But, as we have tried to show, there is no virtue in this for fiction. Not only do the nominalia of fiction not trade blows with the denizens of the world, there is absolutely no predictive gain in integrating their truths. What is wanted is a form of stipulationism in which the truths of fiction leave no metaphysical footprints and no predictive ones either.

The S-model provides that the objects of fiction are a *façon de parler*. Given the way that natural languages are constructed, given that they deeply embed the realist stance, it is only natural that we see our relationships with Holmes and his ilk on the model of our relationships with Saul A. Kripke and his ilk. Perhaps this is a place at which to revisit the previously derided idea of fiction as pretense, although not with a view to recanting my earlier criticism. In our untutored and pre-theoretical moments we report our experience of Holmes objectually. But no one is in any doubt that such reports are a *façon de parler*. Readers of “The Hound of the Baskervilles” do not lose sight of the fact that none of this actually happened. And yet, in some cases — looking for it in the Doyle oeuvre might not be the best place — we are deeply moved by what we read. This is *not* a *façon de parler*. We are not pretending to be deeply moved, or playing at it or making-believe that we are. What is remarkable is that knowing that the denotation of “Ophelia” is a *façon de parler* we are greatly moved by her death. It is a trick, of course, a lovely gift of a trick. We are able to be moved by what we know isn’t there. Our understanding of such things is double-aspected.[60]

The proposition that we are able to be moved by what we know isn’t there gives great offence to philosophers of many stripes, who are moved to back-pedal with a certain earnestness. So we postulate (or borrow) nonexistent but otherwise wholly legitimate objects — the objects of Meinong’s jungle, the objects of possible worlds, the objects of fictional worlds the objects of idealist stipulations. At every turn, these are rehabilitations induced by the objectural biases in-built to language and experience. By and large, these are biases that serve us well, but in the matter of our experience of fiction they lead us into error.

There are logicians galore for whom this is the right way of proceeding. But it is not the right way of proceeding for fiction, as I have been trying to show. Truth in fiction is not a matter of relations between fictional sentences and various kinds of set-theoretic structures (jungles, worlds, whatever). Truth in fiction is a matter entirely of membership in M. Again, people will complain. “That is not real truth”, they will say. They are entirely right. It is truth in fiction.

---

[60] An astute analysis of affective responses to fiction is [Alward, forthcoming, chapter 2].

In recommending an instrumental stipulationism for literary fiction, we make a large investment in fiction’s parasitical nature. Perhaps it is natural to read axiom D as heavily creative. Certainly, it is nothing but right to acknowledge the pivotal role of the author in stocking the M of his story with its enabling fictive sentences. But the great bulk of what is encompassed by M is furnished by the world. What stories are like draws heavily on what the world is like; and the passage of what the world is like into fiction is one in which the fictive operator is picked up in a largely honorific and book-keeping sense. In their stories, authors need the world to be a certain way in order that the action of the story might unfold as the author intends. The default position is that the world in the story is our world, save only as may be required by its fictive sentences. In this regard, understanding a story is like revising a set of beliefs in something like the manner of [Gabbay, Pigozzi and Woods]. Fictive sentences serve both as new inputs and modularity triggers. Someone who understands that *The Hound of the Baskervilles* is a story will not revise his belief set on the strength of the truth of “Sherlock Holmes waved our strange visitor into a chair”, but rather on the strength of that fact that T^F(Sherlock Holmes waved our strange visitor into a chair). But he will produce his version of the maximal account in reverse, with “Sherlock Holmes waved our strange visitor into a chair” making the cut and “T^F(Sherlock Holmes waved our strange visitor into a chair” not. The fact remains that most of what goes into it is already in the world. A great deal of what would occasion the revision of my beliefs also requires membership in a story’s M. When this happens, the author borrows the world for a guest appearance in his story.

Fiction hasn’t the slightest prospect of being integrated with our knowledge of the world. Equipping its stipulations with features designed to ease such integration is worse that a waste of time; is also a mistake. It encourages the idea that, in creating stories, authors create objects. In so saying, we learn a useful lesson about “in”. It is true in Japan that one drives on the left. From this it follows that “One drives on the left” is a sentence contingently true in the world. In the Holmes stories, Holmes is a bit of a prig. But it is not the case that “Holmes is a prig” is contingently true of the world. “Holmes is a prig” is true-in-fiction, but true nowhere. Fiction leaves no metaphysical footprints.

6 ACKNOWLEDGEMENTS

For astute and helpful comments, I am indebted to conversations with Dale Jacquette, Kendall Walton, Nicholas Griffin, Dov Gabbay, Matthew McKeon, Hartley Slater, Peter Alward, Mohan Matthen and Ori Simchen. I thank the Engineering and Physical Sciences Research Council of the United Kingdom for financial support, and Carol Woods and Jane Spurr for technical assistance.

BIBLIOGRAPHY

[Adams *et al.*, 1997] F. Adams, G. Fuller and R. Stecker. The semantics of fictional names, in *Pacific Philosophical Quarterly*, volume 78, pp. 128-148, 1997.

[Akiba, 2000] K. Akiba. Indefiniteness of mathematical objects, *Philosophia Mathematica* volume 3, pp. 26-46, 2000.

[Alward, to appear] P. Alward. *Attitudes Towards Fiction*, to appear.

[Åqvist, 2003] L. Åqvist. Some remarks on performatives in the law, *Artificial Intelligence and Law*, pp. 105-124, 2003.

[Armour-Garb, 2005] B. Armour-Garb. Wrestling with (and without) dialetheism, *Australasian Journal of Philosophy*, vol., 83, pp. 87-102, 2005.

[Austin, 1975] J.L. Austin. *How to Do Things with Words*, Cambridge, MA: Harvard University Press, 1975.

[Azzouni, 1994] J. Azzouni. *Mathematical Myths, Mathematical Practice*, Cambridge: Cambridge University Press, 1994.

[Azzouni, 2004] J. Azzouni. *Deflating Existential Consequence: A Case for Nominalism*, New York: Oxford University Press, 2004.

[Balaguer, 1996] M. Balaguer. A fictionalist account of the indispensable applications of mathematics, *Philosophical Studies*, volume 83, pp. 291-314, 1996.

[Batens, 2000] D. Batens. A survey of inconsistency-adaptive logics, edited by D. Batens, C. Mortenson, G. Priest and J.P. van Bendegem, *Frontiers of Paraconsistent Logic*, pp. 49-73, Baldock, UK: Research Studies Press, 2000.

[Benacerraf, 1973] P. Benacerraf, Mathematical truth, *Journal of Philosophy*, volume LXX, pp. 661-679, 1973.

[Bencivenga, 1980] E. Bencivenga. Free semantics for definite descriptions, *Logique et Analyse*, volume 92, pp. 393-405, 1980.

[Bencivenga, 1991] E. Bencivenga. Free semantics, in K. Lambert, editor, *Philosophical Applications of Free Logic*, New York: Oxford University Press, pp. 98-111, 1991.

[Blocker, 1974] H.G. Blocker. The truth about fictional entities, *Philosophical Quarterly*, volume 24, pp. 27-36, 1974.

[Brown, 2000] B. Brown. Simple natural deduction for weakly aggregative paraconsistent logics, in Dov M. Gabbay, editor, *Frontiers of Paraconsistent Logics*, Baldock, UK: Research Studies Press, 2000.

[Bunge, 1997] M. Bunge. Moderate mathematical fictionalism, in *Philosophy of Mathematics*, E. Agazzi and G. Darvas, editors, pp. 51-71, Dordrecht and Boston: Kluwer, 1997.

[Byrne, 1993] A. Byrne. Truth in fiction: the story continued, *Australasian Journal of Philosophy*, volume 71, pp. 24-35, 1993.

[Caldon and Ignjatović, 2005] P. Caldon and A. Ignjatović. On mathematical instrumentalism, *The Journal of Symbolic Logic*, pp. 778-794, 2005.

[Carlson and Pelletier, 1995] G. N. Carlson and F. J.Pelletier, editors. *The Generic Book*, Chicago: Chicago University Press, 1995.

[Cartwright, 1960] R. Cartwright, Negative Existentials, *Journal of Philosophy*, volume 57, pp. 629-639, 1960.

[Castaneda, 1979] H.-N. Castãnada. Fiction and reality: Their fundamental connections, in *Formal Semantics and LiteraryTheory,* pp. 31-62, John Woods and Thomas Pavel, editors, a special issue of *Poetics*, volume 8, 1979.

[Coleman, 1973] F.X.J. Coleman. A few observations on fictional discourse, in *Aesthetics and Language*, edited by B.R. Tilghman, pp. 31-42, Lawrence: University of Kansas Press, 1973.

[Crittenden, 1973] C. Crittenden. Thinking about non-being, *Inquiry,* volume 16, pp. 290-312, 1973.

[Currie, 1990] G. Currie. *The Nature of Fiction*, Cambridge: Cambridge University Press, 1990.

[da Costa, 1998] N.C.A. da Costa. Paraconsistent logic, in *Stanisaw Jáskowski Memorial Symposium*, pp. 29-35, Department of Logic, Nicholas Copernicus University of Torun, 1998.

[Davis, 2000] M. Davis. *Engines of Logic: Mathematicians and the Origin of the Computer*, New York: Norton, 2000.

[de Kleer, 1986] J. de Kleer. An assumption-based TMS, *Artificial Intelligence,* 28, pp. 157-200, 1986.

[Deutsch, 1985] H. Deutsch. Fiction and fabrication, *Philosophical Studies*, volume 47, pp. 201-211, 1985,

[Devine, 1974] P.E. Devine. The logic of fiction, *Philosophical Studies*, volume 26, pp. 389-399, 1974.

[Doležel, 1979] L. Doležel. Extensional and intuitional narrative worlds, *Formal Semantics and Literary Theory*, pp. 193-211, editors, John Woods and Thomas Pavel, a special of *Poetics*, volume 8, 1979.

[Donnellan, 1966] K. Donnellan. Reference and definite descriptions, *Philosophical Review*, volume 77, pp. 281-304, 1966.

[Conan Doyle, 1981] A. Conan Doyle. *The Hound of the Baskervilles*, London: Penguin Books. Originally published in 1902 by George Newnes, 1981.

[Everett and Hofweber, 2000] A. Everett and T. Hofweber. editors, *Empty Names: Fiction and the Puzzles of Non-existence*, Stanford: CSLI Publications, 2000.

[Field, 1980] H. Field. *Science Without Numbers: A Defense of Nominalism*, Princeton: Princeton University Press, 1980.

[Frege, 1952] G. Frege. On sense and reference, *Translations from the Writings of Gottlob Frege*, pp. 56-78, Peter T. Geach and Max Black, editors, Oxford: Blackwell, 1952.

[Frege, 1967] G. Frege. Begriffsschrift ein der arithmeteschen nachgehildete Formelsprache des reinen Denkins, in *From Grege to Gödel,* Jean van Heijenoort, editor, Cambridge, MA: Harvard University Press, 1967.

[Gabbay, 1996] D. M. Gabbay. *Labelled Deductive Systems*, Oxford: Oxford University Press, 1996.

[Gabbay and Woods, 2003] D. M. Gabbay and John Woods. *Agenda Relevance: A Study in Formal Pragmatics*, Amsterdam: North-Holland, 2003.

[Gabbay and Woods, 2005] D. M. Gabbay and John Woods. *The Reach of Abduction: Insight and Trial*, Amsterdam: Elsevier, 2005.

[Gabriel, 1975] G. Gabriel. *Fiktion und Warheit*, Stuttgart: Fromman-Holzboog, 1975.

[Gabriel, 1979] G. Gabriel. Fiction — a semantic approach, in *Formal Semantics and Literary Theory*, John Woods and Thomas G. Pavel, editors, pp. 245-255, a special issue of *Poetics*, volume 8, Amsterdam: North-Holland, 1979.

[Gale, 1971] R.M. Gale. The fictive use of language, *Philosophy*, volume 46, pp. 324-340, 1971.

[Grice, 1989] P. Grice. *Studies in the Way of Words*, Cambridge, Mass: Harvard University Press, 1989.

[Griffin, 2005] N. Griffin. Through the Woods to Meinong's jungle, in *Mistakes of Reason: Essays in Honour of John Woods,* edited by Kent A. Peacock and Andrew D. Irvine, pp. 15-32, Toronto: University of Toronto Press, to appear in 2005.

[Hailperin and Leblanc, 1959] T. Hailperin and H. Leblanc. Nondesignating singular terms, *The Philosophical Review*, volume 68, pp. 239-243, 1959.

[Hanley, 2004] R. Hanley. As good as it gets: Lewis on truth in fiction, in *Lewisean Themes: The Philosophy of David K. Lewis*, pp. 113-129, Frank Jackson and Graham Priest, editors, New York: Oxford University Press, 2004.

[Heintz, 1979] J. Heintz. Reference and inference in fiction, pp. 85-89, in John Woods and Thomas Pavel, editors, *Formal Semantics and Literary Theory*, a special issue of *Poetics*, volume 8, 1979.

[Hintikka, 1959] J. Hintikka. Existential presuppositions and existential commitments, *The Journal of Philosophy*, pp. 125-137, volume 56, 1959.

[Hodes, 1984] H. T. Hodes. Logicism and the ontological commitments of arithmetic, *Journal of Philosophy*, pp. 123-149, volume 81, 1984.

[Hoffman, 1999] S. Hoffman. *Mathematics as Makebelieve: A Constructive Empiricist Account*, Ph.D Thesis, University of Alberta, 1999.

[Howell, 1974] R. Howell. The logical structure of pictorial representation, *Theoria*, pp. 76-109, volume 40, 1974.

[Howell, 1976] R. Howell. Review of John Woods: *The Logic of Fiction, The Journal of Aesthetics and Art Criticism*, pp. 354-355, volume 34, 1976.

[Howell, 1979] R. Howell. Fictional objects: How they are and how they aren't, in *Formal Semantics and Literary Theory,* pp. 50-72, John Woods and Thomas Pavel, editors, a special issue of *Poetics*, volume 8, Amsterdam: North-Holland, 1979.

[Howell, 1983] R. Howell. Review of *Non-Existent Objects, Journal of Philosophy*, pp. 163-173, volume 80,1983.

[Ingarden, 1931] R. Ingarden. *Das literarische Kunstwerk. Eine Untersuchung aus dem Grenzgebiet der Ontologie, Logik und Literaturwissenschaft*, Halle: Neimeyer, 1931.

[Jacquette, 1996] D. Jacquette. *Meinongian Logic: The Semantics of Existence and Nonexistence*, Berlin: Walter de Gruyter, 1996.

[Jacquette, 1998] D. Jacquette. Paraconsistent logical consequence, *Journal of Applied Non-Classical Logics*, volume 8, pp. 337-351, 1998.

[Jacquette, 2001] D. Jacquette. Truth and fiction in David Lewis' critique of Meinongian semantics, *Metaphysica*, volume 2, pp. 73-106, 2001.

[Jacquette, to appear] D. Jacquette. Logic for Meinongian object theory semantics, in Dov M. Gabbay and John Woods, editors, *Handbook of the History of Logic*, volume 5, *Logic From Russell to Gentzen*, Amsterdam: Elsevier, to appear in 2006.

[Kaplan, 1973] D. Kaplan. Bob and Carol and Ted and Alice, in *Approaches to Natural Language*, edited by Jaakko Hintikka, J.M. Moravcsik and Patrick Suppes, pp. 490-518, Dordrecht: Reidel, 1973.

[Kivy, 1993] P. Kivy. *The Fine Art of Repetition: Essays in the Philosophy of Music*, New York: Cambridge University Press, 1993.

[Körner, 1967] S. Körner. On the relevance of post-Gödelian mathematics to philosophy, in *Problems in the Philosophy of Mathematics*, Imre Lakatos, editor, pp. 133-137, Amsterdam: North-Holland, 1967.

[Kripke, 1972] S. Kripke. Naming and necessity, in *Semantics and Natural Language*, edited by Donald Davidson and Gilbert Harman, pp. 253-355, 763-769, Dordrecht: Reidel, 1972.

[Kripke, 1973] S. Kripke. *Reference and Existence*, unpublished typescript. Presented as the John Locke Lectures, Oxford University, 1973.

[Kripke, 1979] S. Kripke. Speaker's reference and semantic reference, in *Contemporary Perspectives in the Philosophy of Language*, edited by Peter A. French, Theodore E. Uehling Jr., and Howard K. Wettstein, pp. 6-27, Minneapolis: University of Minnesota Press, 1979.

[Kripke, 1982] S. Kripke. *Wittgenstein's Private Language Argument*, Oxford: Blackwell, 1982.

[Landman, 1986] F.W. Landman. *Towards a Theory of Information: The Status of Partial Objects in Semantics*, Dordrecht: Foris, 1986.

[Larmarque and Olsen, 1994] P. Larmarque and S. Olsen. *Truth, Fiction and Literature, A Philosophical Perspective*, Oxford: Oxford University Press, 1994.

[Lambert, 1963a] K. Lambert. Notes on E!III: A theory of descriptions, *Philosophical Studies*, volume 13, pp. 51-59, 1963a.

[Lambert, 1963b] K. Lambert. Existential import revisited, *Notre Dame Journal of Formal Logic*, volume 4, pp. 288-292, 1963b.

[Lambert, 1964] K. Lambert. Notes on E!IV: A reduction in free quantification theory with identity and descriptions, *Philosophical Studies*, volume 15, pp. 85-88, 1964.

[le Poidevin, 1995] R. Le Poidevin. Worlds within worlds? The paradoxes of embedded fiction, *British Journal of Aesthetics*, volume 35, pp. 227-238, 1995.

[Lear, 1982] J. Lear. Aristotle's philosophy of mathematics, *Philosophical Review,* volume 91, pp. 161-192, 1982.

[Lejewski, 1954] C. Lejewski. Logic and existence, *British Journal for the Philosophy of Science*, volume 5, pp. 104-119, 1954.

[Levinson, 2000] S. C. Levinson. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, Cambridge, MA: MIT Press, 2000.

[Lewis, 1978] D. Lewis. Truth in fiction, *American Philosophical Quarterly*, volume 15, pp. 37-46, 1978.

[Lewis, 1983] D. Lewis. *Philosophical Papers*, volume 1, Oxford: Oxford University Press, 1983.

[Maddy, 2005] P. Maddy, Mathematical existence, *The Bulletin of Symbolic Logic*, volume 11, pp. 351-376, 2005.

[Martin and Schotch, 1974] R. Martin and P. Schotch. The meaning of fictional names, *Philosophical Studies*, volume 26, pp. 377-388, 1974.

[McKeon, 2005] M. McKeon. On the substitutional approach to logical consequence, in *Mistakes of Reason: Essays in Honour of John Woods,* edited by Kent A,. Peacock and Andrew D. Irvine, Toronto: University of Toronto Press, to appear in 2005.

[Meheus *et al.*, forthcoming] J. Meheus, L. Verhoeven, M. Van Dyck and D. Provijn. Ampliative adaptive logics and the foundation of logic-based approaches to abduction, in *Logical and Computational Aspects of Model-Based Reasoning,* Lorenzo Magnani, Nancy J. Nersessian and Claudio Puzzi, editors, Dordrecht and Boston: Kluwer, forthcoming.

[Meinong, 1915] A. Meinong. *Über Möglichkeit und Warhscheinlichkeit*, Leipzig: Barth, 1915.

[Mooij, 1993] J.J.A. Mooij. *Fictional Realities: The Uses of Literary Imagination*, Amsterdam and Philadelphia: John Benjamins, 1993.

[Nash, 1950] J. Nash. Equilibrium, in the *Proceedings of the National Academy of Sciences*, volume 36, pp. 48-49, 1950.

[Neale, 1990] S. Neale. *Descriptions*, Cambridge, MA: MIT Press, 1990.

[Oversteegen, 1982] J.J. Oversteegen. *Beperkingen, Methodlogische recepten en andere voorondelen in de moderne literatuurwetenschap*, Utrecht: HES Uitgevers, 1982.

[Parsons, 1975] T. Parsons. A Meinongian analysis of fictional objects, *Grazer Philosophische Studien*, volume 1, pp. 73-86, 1975.

[Parsons, 1978] T. Parsons. Review of J. Woods' *The Logic of Fiction, Synthese,* volume 39, pp. 155-164, 1978.

[Parsons, 1980a] T. Parsons. *Non-Existent Objects*, New Haven, CT: Yale University Press, 1980a.

[Parsons, 1980b] T. Parsons. A prolegomenon to Meinongian semantics, *Journal of Philosophy*, volume 77, pp. 561-586, 1980b.

[Parsons, 1982] C. Parsons. Objects and logic, *The Monist*, volume 65, pp. 491-516, 1982.

[Pavel, 1975] T. Pavel. 'Possible worlds' in literary semantics, *The Journal of Aesthetics and Art Criticism*, volume 34, pp. 165-176, 1975.

[Pavel, 1986] T. Pavel. *Fictional Worlds*, Cambridge MA: Harvard University Press, 1986.

[Phillips, 1999] J. F. Phillips. Truth and inference in fiction, *Philosophical Studies*, volume 94, pp. 273-293, 1999.

[Plantinga, 1974] A. Plantinga. *The Nature of Necessity*, Oxford: Oxford University Press, 1974.

[Priest, 1979] G. Priest. The logic of paradox, *Journal of Philosophical Logic*, volume 8, pp. 219-241, 1979.

[Priest, 1987] G. Priest. *In Contradiction: A Study of the Transconsistent*, Dordrecht: Kluwer Academic, 1987.

[Priest, 1997] G. Priest. Sylvan's box: a short story and ten morals, *Notre Dame Journal of Formal Logic*, volume 38, pp. 573-582, 1997.

[Priest, 2001] G. Priest. *An Introduction to Non-Classical Logic*, Cambridge: Cambridge University Press, 2001.

[Quine, 1939] W.V. Quine. Designation and existence, *The Journal of Philosophy*, volume 36, pp. 701-709, 1939.

[Radford, 1975] C. Radford. How can we be moved by the fate of Anna Karenina? I, *The Aristotelian Society*, volume XLIX, pp. 67-80, Supplementary 1975.

[Rapaport, 1978] W. J. Rapaport. Meinongian theories and a Russellian paradox, *Noûs,* 12, pp. 153-158, 1978.

[Recanati, 1993] F. Recanati. *Direct Reference*, Oxford: Blackwell, 1993.

[Resnik, 1997] M. Resnik. *Mathematics as a Science of Patterns*, Oxford: Clarendon Press, 1997.

[Routley, 1966] R. Routley. Some things do not exist, *Notre Dame Journal of Formal Logic* volume 7, pp. 251-276, 1966.

[Routley and Routley, 1973] R. Routley and V. Routley. Rehabilitating Meinong's theory of objects, *Revue Internationale de Philosophie*, volume 27, pp. 224-254, 1973.

[Routley and Meyer, 1976] R. Routley and R. K. Meyer. Dialectical logic, classical logic and the consistency of the world, *Studies in Soviet Thought*, vol. 16, pp. 1-25, 1976.

[Routley, 1979] R. Routley. The semantic structure of fictional discourse, in *Formal Semantics and Literary Theory,* pp. 3-30, John Woods and Thomas Pavel, editors, a special issue of *Poetics*, volume 8, 1979.

[Routley, 1981] R. Routley. *Exploring Meinong's Jungle and Beyond*: *An Investigation of Noneism and the Theory of Items*, Canberra: Australian National University, 1981.

[Russell, 1937] B. Russell. *The Principles of Mathematics*, London: George Allen and Unwin, 1903. Second edition published in 1937.

[Salmon, 1998] N. Salmon. Nonexistence, *Noûs*, volume 32, pp. 277-319, 1998.

[Schotch and Jennings, 1989] P. Schotch and R. Jennings. On detonating, in Graham Priest, Richard Routley, Jean Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*, pp. 306-327, Munich: Philosophia Verlag, 1989.

[Scruton, 1974] R. Scruton. *Art and Imagination: A Study in the Philosophy of Mind*, New York: Barnes and Noble, 1974.

[Searle, 1970] J. Searle. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press, 1970.

[Searle, 1975] J. Searle. The logical status of fictional discourse, *New Literary History*, volume 6 pp. 319-332, 1975.

[Simons, 1999] P. Simons. On what there isn't: The Meinong-Russell dispute, in *Bertrand Russell: Critical Assessments,* Andrew Irvine, editor, volume III of *Language, Knowledge and the World*, pp. 69-100, London: Routledge, 1999.

[Slater, 1987] H. Slater. Fictions, *British Journal of Aesthetics*, volume 27, pp. 143-155, 1987.

[Slater, 1993] H. Slater. The incoherence of the aesthetic response, *British Journal of Aesthetics*, volume 33, pp. 168-172, 1993.

[Slater, 2002] H. Slater. *Logic Reformed*, Bern: Peter Lang, 2002.

[Smiley, 1960] T. Smiley. Sense without denotation, *Analysis*, volume 20, pp. 125-135, 1960.

[Herrnstein Smith, 1978] B. Herrnstein Smith. *On the Margins of Discourse*, Chicago: University of Chicago Press, 1978.

[Stone, 1937] W.H. Stone. Applications of the theory of Boolean rings to general topology, *Transactions of the American Mathematical Society*, volume 41, pp. 375-481, 1937.

[Stove, 1999] D. Stove. Cole Porterism: The jazz age in the philosophy of science, *Against the Idols of the Age*, edited by Roger Kimball, pp. 3-32 New Brunswick NJ: Transaction Publishers, 1999.

[Strawson, 1950] P.F. Strawson. On referring, *Mind*, volume 59, pp. 320-344, 1950.

[Suppes, 1962] P. Suppes. Models of data, in Ernest Nagel, Patrick Suppes and Alfred Tarski (eds.), *Logic, Methodology and Philosophy of Science*, Stanford: Stanford University Press pp. 287-301, 1962.

[Tait, 1981] W.A. Tait. Finitism, *Journal of Philosophy*, volume 78, pp. 524-546, 1981.

[Taylor, 2000] K. Taylor. Emptiness without compromise, in *Empty Names: Fictions and the Puzzles on Non-existence*, edited by Anthony Everett and Thomas Hofweber, pp. 17-36, Stanford: CSLI Publications, 2000.

[Tharp, 1989] L. Tharp. Myth and mathematics: A conceptualistic philosophy of mathematics I, *Synthese,* volume 81, pp. 167-201, 1989.

[Thomas, 2000] R.S.D. Thomas. Mathematics and fiction I: Identification, *Logique et Analyse*, volume 43, pp. 301-340, 2000.

[Thomas, 2005] R.S.D. Thomas. Mathematics and fiction II: Analogy, *Logique et Analyse*, volume 45, pp. 185-228, 2005.

[Thomasson, 1998] A. Thomasson. *Fiction and Metaphysics*, Cambridge: Cambridge University Press 1998.

[Tiles, 1988] M. Tiles. Scientific dreamspace: Symbolic forms and scientific theories, *International Studies in Philosophy of Science*, volume 2, pp. 189-204, 1988.

[Torretti, 1981] R. Torretti. Three kinds of mathematical fictionalism, in *Scientific Philosophy*, Joseph Agassi and Robert S. Cohen, editors, pp. 399-414, Dordrecht and Boston: Reidel, 1981.

[Toulmin, 1953] S. Toulmin. *The Philosophy of Science: An Introduction*, London: Hutchinson, 1953.

[Urmson, 1976] J.O. Urmson. Fiction, *American Philosophical Quarterly*, volume 13, pp. 153-157, 1976.

[Vaihinger, 1935] H. Vaihinger. *The Philosophy of 'As If': A System of the Theoretical Practical and Religious Fictions of Mankind*, 2^{nd} edition. Translated by C.K. Ogden, London: Routledge and Kegan Paul, 1935.

[van Fraassen, 1966a] B. van Fraassen. The completeness of free logic, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, volume 12, pp. 219-234, 1966a.

[van Fraassen, 1966b] B. van Fraassen. Singular terms, truth-value gaps, and free logic, *The Journal of Philosophy*, volume 63, pp. 481-494, 1966b.

[van Inwagen, 2000] P. van Inwagen. Quantification and fictional discourse, *Empty Names: Fiction and the Puzzles of Non-existence*, Anthony Everett and Thomas Hofweber, editors, pp. 235-247, Stanford: CSLI Publications 2000.

[van Neumann and Morgensern, 1944] J. van Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*, Princeton: Princeton University Press, 1944.

[von Feytag-Löringhoff, 1951] B. von Freytag-Löringhoff. *Philosophical Problems of Mathematics*, translated by Amethe von Zeepelin, New York: Philosophical Library, 1951.

[Wagner, 1982] S. Wagner. Arithmetical fiction, *Pacific Philosophical Quarterly*, volume 63, pp. 255-269, 1982.

[Walton, 1987a] K. Walton. How remote are fictional worlds from the real world?, *The Journal of Aesthetics and Art Criticism*, volume 37, pp. 11-23, 1978a.

[Walton, 1978b] K. Walton. On fearing fictions, *Journal of Philosophy*, volume 75, pp. 5-27, 1978b.

[Walton, 1990] K. Walton. *Mimesis as Make-Believe*, Cambridge MA: Harvard University Press, 1990.

[Wang, 1986] H. Wang. *Beyond Analytic Philosophy*, Cambridge, MA: MIT Press, 1986.

[Wittgenstein, 1958] L. Wittgenstein. *Philosophical Investigations,* G.E.M. Anscombe, translator, New York: Macmillan, 1958.

[Woltersdorff, 1976] N. Woltersdorff. Worlds of works of art, *The Journal of Aesthetics and Art Criticism*, volume 35, pp. 121-142, 1976.

[Woods, 1969] J. Woods. Fictionality and the logic of relations, *The Southern Journal of Philosophy*, volume 7, pp. 51-64, 1969.

[Woods, 1974] J. Woods. *The Logic of Fiction: Philosophical Soundings of Deviant Logic*, The Hague and Paris: Mouton, 1974.

[Woods and Pavel, 1979] J. Woods and T. G. Pavel. editors, *Formal Semantics and Literary Theory*. A special issue of *Poetics,* volume 8, Amsterdam: North-Holland, 1979.

[Woods, 1986] J. Woods. God, genidentity and existential parity, *Grazer Philosophische Studien,* 25-26, pp. 181-196, 1986.

[Woods, 2005] J. Woods. Dialectical considerations on the logic of contradiction I, *Logic Journal of the IGPL*, volume 8, pp. 231-260, 2005.

[Woods, forthcoming] J. Woods. Probability in the law, forthcoming.

[Zalta, 1983] E. N. Zalta. *Abstract Objects: An Introduction to Axiomatic Metaphysics*, Dordrecht and Boston: Reidel, 1983.

[Zalta, 2000] E. N. Zalta. The road between pretense theory and abstract object theory, in *Empty Names: Fiction and the Puzzles of Non-existence*, Anthony Everett and Thomas Hofweber, editors, pp. 117-147, Stanford: CSLI Publications, 2000.

COUNTERFACTUALS, CAUSATION, AND PREEMPTION

John Collins

1 INTRODUCTION

A *counterfactual* is a conditional statement in the subjunctive mood. For example:

> If Suzy hadn't thrown the rock, then the bottle wouldn't have shattered.

The philosophical importance of counterfactuals stems from the fact that they seem to be closely connected to the concept of *causation*. Thus it seems that the truth of the above conditional is just what is required for Suzy's throw to count as a cause of the bottle's shattering. If philosophers were reluctant to exploit this idea prior to 1970, it was because of a widespread feeling that the truth-conditions of the counterfactual conditional were not sufficiently well understood. The development of a formal semantics for counterfactuals by Robert Stalnaker [1968] and David Lewis [1973b] stands as a major recent achievement in philosophical logic.

§2 presents the standard Stalnaker-Lewis semantics for the counterfactual conditional and develops some of the logical features of counterfactuals. §3 presents Lewis's original counterfactual theory of causation [Lewis, 1973a], and explains the problems that eventually led him to abandon the theory in its original form. The remainder of the article surveys the current state of counterfactual theories of causation, by presenting, in §§4, 5, and 6, three recent contending accounts, due to Lewis [2000; 2004a], Yablo [2002; 2004] and Hall [2004a; 2004b].

The discussion does not aim to be exhaustive, but focuses on the central issue of *preemption*, which has proved to be the major hurdle for counterfactual theories. For a more complete picture, the reader is referred to the papers in the collection edited by Collins, Hall, and Paul [2004a]. The present article aims to provide an brief introduction to this currently very lively area of applied philosophical logic. For a more comprehensive introductory survey, Hall and Paul [forthcoming] is also highly recommended.

The contemporary literature on counterfactuals and causation includes a vast and potentially bewildering collection of examples and counterexamples. The present article focuses on six central types of example, which are labeled (E1)–(E6) so that the reader may more easily distinguish them from other less important particular cases mentioned in passing.

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

2 THE LOGIC OF COUNTERFACTUALS

The counterfactual 'If A were true, then C would be true' with antecedent A and consequent C is sometimes written '$A \mathbin{\Box\!\!\rightarrow} C$' in order to distinguish it from other kinds of conditional statement. For example the counterfactual must be distinguished from both the material conditional of first-order logic, and the "strict conditional" of entailment. The truth-functional material conditional '$A \rightarrow C$' is logically equivalent to '$\sim A \vee C$' and thus has truth-conditions weaker than those of the corresponding counterfactual conditional. Not every counterfactual with a false antecedent or a true consequent is true. The strict conditional, on the other hand, is too strong; it might be true that:

> If this match had been struck, it would have lighted.

but the lighting of the match is not logically entailed by its being struck.

A correct account of the semantics of the counterfactual will then, presumably, locate it somewhere between these two extremes. But where? One obvious thought is that '$A \mathbin{\Box\!\!\rightarrow} C$' is true if and only if C is entailed not by A alone, but by A in conjunction with certain other truths, including, perhaps, the laws of nature. Thus it might well be that the lighting of the match *is* entailed by its being struck, in conjunction with the laws of nature, and certain other true matters of fact — for example the presence of sufficient oxygen in the atmosphere.

The problem with this idea is that there is no single fixed set of truths that will do the job for all A and C. That is because counterfactuals are *nonmonotonic*. In other words the inference pattern:

$$\begin{array}{rc} & A \mathbin{\Box\!\!\rightarrow} C \\ \cline{2-2} \text{so:} & (A\&B) \mathbin{\Box\!\!\rightarrow} C \end{array}$$

is invalid. It may be true, for example, that the match would light if struck, and yet not true that it would light if struck *in the absence of oxygen*.

But the counterfactual conditional will only serve as a fit tool for philosophical analysis if we have a firm grasp of its logic and truth-conditions. In a famous essay critical of philosophical use of the counterfactual idiom, Nelson Goodman framed the challenge this way. In evaluating the truth of the counterfactual '$A \mathbin{\Box\!\!\rightarrow} C$' we want to hold fixed all those truths that are "cotenable" with the truth of the antecedent A. Yet what might it mean for a proposition to be cotenable with A other than that the proposition is one that would still be true even if A were true? If one hopes to provide a semantics for counterfactuals in terms of what is cotenable with a given antecedent, then one had better not rest this on a counterfactual analysis of cotenability [Goodman, 1947].

It was the work of Stalnaker [1968] and Lewis [1973], which in turn grew out of the development of possible world semantics for modal logic in the 1960s, that first convinced some of the skeptics that counterfactual conditionals were indeed philosophically respectable.

The Stalnaker-Lewis approach starts from the assumption that the set of all possible worlds may be weakly ordered with respect to the "comparative similarity" of those world to the actual world. Since this ordering is transitive and connected, it is useful heuristically to think of it as a comparative "closeness" relation. If A is any proposition, call a world at which A is true an *A-world.* Then, in Lewis's formulation, the truth-condition for the counterfactual may be stated in this way:

> '$A \mathrel{\Box\!\!\rightarrow} C$' is true if and only if some $(A\&C)$-world is more similar to the actual world than any $(A\& \sim C)$-world is.

If we simplify things by assuming that for each A there is always a single A-world most similar to the actual world, then the condition becomes:

> '$A \mathrel{\Box\!\!\rightarrow} C$' is true if and only if C is true at the most similar A-world to the actual world.

Now similarity is itself a philosophically problematic concept, but even without settling on any particular criteria for making comparative judgments, we can see how much of the logic of counterfactuals follows simply from the fact that any candidate similarity ordering, like a closeness ordering, must be connected and transitive.

We can see, for example, why strengthening the antecedent of a counterfactual may lead from truth to falsehood. The fallacy is akin to that made by a person who infers from the fact that there is no bank in the closest town to here, that there is no bank in the closest town to here with a restaurant. We can also see immediately that contraposition fails for the counterfactual conditional, i.e. that the inference

$$\begin{array}{rc} & A \mathrel{\Box\!\!\rightarrow} C \\ \hline \text{so:} & \sim C \mathrel{\Box\!\!\rightarrow} \sim A \end{array}$$

is fallacious. It doesn't follow from the fact that the closest town with a restaurant has no bank that the closest town with a bank has no restaurant.

A similar exploitation of the analogy with spatial closeness will convince us that counterfactuals are not transitive. That is, we can add to our list of counterfactual fallacies the following:

$$\begin{array}{rc} & A \mathrel{\Box\!\!\rightarrow} B \\ & B \mathrel{\Box\!\!\rightarrow} C \\ \hline \text{so:} & A \mathrel{\Box\!\!\rightarrow} C \end{array}$$

Moving beyond these purely logical points, however, more will need to be said about the comparative similarity relation. The problem is that similarity admits of different respects. One possible state of affairs may be more similar to actuality than another in one respect, and less similar to actuality in another. Often it is the context of utterance that determines the particular similarity ordering that a speaker has in mind. But insofar as we want counterfactuals to provide objective truth-conditions for causal statements, this question of context-dependence is fairly pressing. How are the various respects of similarity to be weighed in order to enable a single overall comparison?

It seems clear that similarities with respect to the laws of nature should generally outweigh similarities with respect to accidental matters of fact. But this cannot be an invariable rule. For if, as we are assuming, the laws of nature are deterministic in both temporal directions, any supposed change in the way things presently are, will be propagated, via the laws, into a divergent past as well as a divergent future. Matching the past history of the actual world is very important for similarity, it appears, even if the match can be obtained only by allowing a minor localized violation of the laws of nature (a "small miracle").

This is not to deny that we *are* sometimes prepared to speak as though things would have been different in the past had they been different now. For example I might say: "If I had jumped out the window just now, there would have to have been a safety net in place, I'm not crazy!" But note the "have to have been" construction in that sentence that serves as a syntactic indicator of the appropriateness of the *backtracking* interpretation (see [Lewis, 1979]). The point is simply that this backtracking interpretation is nonstandard. Causes always, or at least typically, precede their effects. Surely this asymmetry should be reflected by a corresponding asymmetry in the counterfactuals. But since the temporal asymmetry of causation seems to be a merely contingent feature of the actual world, it seems wrong to build this asymmetry into the analysis of counterfactuals by stipulation. Lewis [1979] pursues the more ambitious goal of identifying criteria for a comparative similarity relation that rule out backtracking in worlds like the actual world, but without making backward causation an *a priori* impossibility.

Here is the rough idea. Suppose that c is some event that actually occurred at time t, and consider a counterfactual whose antecedent asks us to suppose that c hadn't occurred. A non-backtracking reading of this counterfactual will be ensured provided that all the closest possible worlds to the actual world at which c doesn't occur are worlds in which (i) past history up until some point shortly before time t perfectly matches the history of the actual world, and (ii) this perfect match results from a small "divergence" miracle. And this is correct, Lewis suggests, because our world happens to be such that there is no possible world w in which (i) c fails to occur, (ii) the future of w *after* time t exactly matches the actual future, and (iii) this match results from a small "reconvergence" miracle.

A important recent paper by Adam Elga has raised a serious difficulty for Lewis's attempt to rule out backtracking contingently. Elga argues that statistical mechanics provides examples which demonstrate that reconvergence to the actual world requires no greater violation of the laws than divergence from it does. See Elga [2000], Albert and Loewer [2005].

3 THE COUNTERFACTUAL THEORY OF CAUSATION

An early statement of a counterfactual analysis of causation can be found in the work of David Hume [1748, Section VII], where he writes:

> we may define a cause to be *an object followed by another, and where all the objects, similar to the first are followed by objects similar to the second.* Or, in other words, *where, if the first object had not been, the second never had existed.*

In fact, of course, the passage just quoted contains two quite different accounts of the causal relation. While the second sentence treats causation in counterfactual terms, the first expresses an idea that is closer to what has become known as a *regularity theory* of causation.

Regularity theories tended to be the more favored of these two broad approaches prior to the development of the Stalnaker-Lewis semantics for the counterfactual. See, e.g. Mackie [1965]. A regularity analysis of causation states, roughly, that:

> One event *c* is a cause of another event *e* just in case *c* is a member of some minimal set of actual events that are jointly sufficient, given the laws of nature, for the occurrence of the effect.

But regularity theories of causation have always faced difficulties. One is the problem of distinguishing cause from effect. For if *e* is caused by *c*, there will often be a set of conditions including *e* which, in conjunction with the laws, entail that *c* occurs.

Another difficulty is the problem of distinguishing genuine causes from inefficacious epiphenomenal by-products of a causal process. If *d* and *e* are independent effects of some common cause *c*, then it may well be the case that *d* belongs to a minimal set of conditions which, along with the laws, are sufficient for *e*. Then *d*, which is a epiphenomenal by-product of the process by which *c* caused *e*, will be incorrectly counted as a cause of *e*.

In his seminal article [1973a], Lewis pointed out that a counterfactual analysis of the causal relation is at a distinct advantage over a regularity theory in both of these cases. Given an appropriate ban on a backtracking interpretation of the relevant counterfactuals, one may simply deny the truth of the counterfactuals that might otherwise cause problems. Suppose that *e* wouldn't have occurred if *c* hadn't occurred. Might it also be the case that if *e* hadn't occurred then *c* wouldn't have? Not unless the latter conditional is understood in the backtracking sense. Similarly, if *d* and *e* are independent effects of a common cause *c*, then the only reason one might be tempted to believe that *if d hadn't occurred then e wouldn't have occurred* is if one thinks that *if d hadn't occurred then neither would c*. Once again, this would involve an illicit backtracking reading of the latter conditional.

Say that one event *e* is *counterfactually dependent* on another event *c* when *e* wouldn't have occurred if *c* hadn't. The simplest counterfactual analysis of causation would simply construe causation as counterfactual dependence.

But this simplest account cannot be correct, as is demonstrated by the following example:

(E1) **Early Preemption** Suzy throws a rock at a bottle. The rock hits the bottle and shatters it. Billy was standing by with a second rock. Had Suzy not thrown her rock, Billy would have shattered the bottle by throwing his rock.

The problem here is that the shattering of the bottle is indeed caused by Suzy's throw, despite the fact that the shattering is not counterfactually dependent on the throw. According to Lewis, the key to reinstating the throw as cause is to recognize that there is a *chain of events* initiated by the throw and terminating in the shattering, such that each event in the chain is counterfactually dependent on the one before it. Call such a chain of events a *chain of counterfactual dependence*. In this case, each event in the chain is the event of Suzy's rock being located in mid-flight at some point between her hand and the bottle. Although Billy was standing by with appropriate intent and deadly aim, no such chain of events connects his standing by with the shattering.

In cases like Early Preemption we have a *chain of counterfactual dependence* leading from a cause to an effect, without the effect being counterfactually dependent on the cause. The problem with preemption, then, seems to stem directly from something we noted in the previous section: the non-transitivity of the counterfactual conditional. While causation appears to be a transitive relation, counterfactual dependence is not.

If this diagnosis of the problem is correct, then the fix is very straightforward. We should simply take causation to be the *ancestral* of the counterfactual dependence relation. Thus we arrive at Lewis's [1973] counterfactual analysis of causation.

(C1) Suppose that c and e are distinct events, and let C and E respectively be the propositions that the events c and e occur. Then e is *counterfactually dependent* on c when the following two counterfactual conditionals are true:

(i) $C \mathrel{\Box\!\!\rightarrow} E$

(ii) $\sim C \mathrel{\Box\!\!\rightarrow} \sim E$

(C2) A *causal chain* is a finite sequence of actual events such that each event in the sequence is counterfactually dependent on the previous event.

(C3) One event is a *cause* of another if and only if there is a causal chain leading from the first to the second.

Note that it is precisely the failure of contraposition for the counterfactual conditional that leaves room for the ban on backtracking, by allowing that e may be counterfactually dependent on c without c also being counterfactually dependent on e.

This original analysis was designed to deal with cases of preemption like (E1). However the phenomenon of preemption has proved to be a far more serious difficulty for the counterfactual theory of causation than Lewis believed it to be in 1973.

The problem is that there are varieties of preemption that cannot be dealt with by the transitivity strategy. Here is one such:

(E2) **Late Preemption** Billy and Suzy both throw rocks at a bottle. Suzy's rock gets there first, hitting the bottle and shattering it. Billy's rock flies through the now empty space where the bottle was standing.

This example differs in one key respect from the previous case. In Late Preemption no stepwise dependent chain of events can be traced from Suzy's throw to the shattering of the bottle. For consider all the events in the process initiated by Suzy's throw prior to the shattering, that is: Suzy's rock being located at various positions at various times in mid-flight. The shattering fails to depend on *any* such event, because, had Suzy's rock failed to be there, the bottle would still have been shattered by Billy's rock.

Here is another way of seeing the difference between the Early and Late Preemption examples. In cases like Early Preemption, the possible process initiated by the preempted standby that would otherwise have led to the effect, is cut off by the process leading from the preempting cause to the effect, at some time *before* the effect occurs. In Late Preemption, this "cutting-off" takes place only when the effect occurs. (This way of looking at things explains the "early/late" terminology.)

Furthermore, as Jonathan Schaffer discovered, there are possible cases of preemption that do not involve any kind of cutting-off at all. In Schaffer's example of "Trumping Preemption" the absence of any cutting, either early or late, is guaranteed simply by stipulating that the causal process in question works by action at a distance.

(E3) **Trumping Preemption** The laws of magic are such that what happens at midnight is determined by the first spell cast the previous day. At noon Merlin casts the first spell of the day: a spell to turn the Prince into a frog. At six that evening Morgana casts a spell to turn the Prince into a frog. At midnight the Prince turns into a frog.

This example also causes problems for Lewis's 1973 theory. The transfiguration of the prince is not counterfactually dependent on Merlin's spell, since if Merlin had not cast the spell, the prince would still have been turned into a frog by Morgana's spell. In addition, there is no chain of counterfactual dependence leading from Merlin's spell to the transfiguration, since the example stipulates that none is required.

It is tempting to dismiss such fantastic cases as being too far-fetched to be relevant to any discussion of causation as it is in the actual world. But this would be too hasty a dismissal. The fanciful nature of Schaffer's story merely helps make the causal structure of the example clear. The important point is that it seems perfectly possible that, for all we know, some actual cases of causation work that way. Since the possibility of causation by trumping preemption cannot be ruled out *a priori*, a theory of causation will only be adequate if it can deal with such cases.

4 CAUSATION AS INFLUENCE

One tempting thought is that the problems facing the counterfactual theory might be solved by taking events to be modally "fragile", i.e. by claiming that any difference in time, or manner of occurrence makes for a numerically distinct event. In our Late Preemption example, the bottle would still have been shattered had Suzy not thrown her rock, but it would have been shattered a moment later by Billy's rock, and, presumably, shattered in a slightly different way. Thus, if the actually occurring shattering is construed as a modally fragile event (i.e. an event with a particularly rich essence) then the counterfactual dependence of the actual shattering on Suzy's throw is restored, since, had Suzy's throw not taken place, the shattering that would have occurred instead would have been a *different* shattering; a numerically distinct event.

This "fragility strategy" is discussed by Lewis in various places (e.g. [1986b, 193–199; 2004a, 85–90]. As Lewis recognizes, there are various reasons for rejecting the approach.

For one thing, the "uncommonly stringent" conditions of occurrence that the fragility strategy imposes are at odds with our ordinary standards of event identity. Lewis points out that we are usually quite happy to allow that one and the same event might have been delayed, as, for example, when a seminar talk is postponed rather than cancelled [2004a, 86].

Secondly, the fragility strategy produces spurious cases of causation, as in Lewis's example of the "Poison and the Pudding". Suppose a poison kills its victim more slowly and painfully when taken on a full stomach. The victim eats some pudding and then drinks the poison. If the victim's actual death is construed as modally fragile, then it is an event that would not have occurred had the pudding not been eaten. Yet the eating of the pudding was not a cause of his death. [Lewis 1986b, 198-199].

Still, the central idea of fragility has a role to play in Lewis's [2000; 2004a] revised account of "causation as influence". While remaining neutral about the ordinary identity conditions for events, Lewis proposes that we introduce a new technical term to refer to events *construed as* modally fragile. Say that an *alteration* of an event is either the very fragile version of the event that actually occurs, or a fragile alternative to it that differs slightly with respect to time or manner of occurrence. Lewis then suggests that we "look at the pattern of counterfactual dependence of alterations of the effect upon alterations of the cause". Say that event c *influences* event e when there is a substantial range $c_1, c_2, c_3, \ldots$ of alterations of c, and a substantial range $e_1, e_2, e_3, \ldots$ of alterations of e, such that if c_1 had occurred then e_1 would have occurred, and if c_2 had occurred then e_2 would have occurred, and so on [Lewis, 2004a, 91]. Our original notion of counterfactual dependence was a notion of "whether–whether" dependence. One event is dependent on another is this sense just in case whether or not the event occurs depends on whether or not the other occurs. But there are other varieties of dependence. Lewis's idea is that we should think of degree of causal influence as being determined by the extent to which *whether*, *when*, and *how* one thing happens depends counterfactually on whether, when, and how, something else happens. Hall and Paul have usefully labeled this idea the "counterfactual covariation" theory of causation [Hall and Paul, forthcoming)]

As in Lewis's original counterfactual theory, causation is now taken to be the ancestral of the influence relation. One event c causes another event e if and only if there is a chain of stepwise influence from c to e.

Let's see how this idea works in our case of Late Preemption. If we consider small alterations to the time of Suzy's throw, we obtain corresponding alterations to the time of shattering (provided of course that the throw is not so much delayed that Billy's rock gets there first). Or consider small alterations to the manner in which Suzy throws her rock; alterations, perhaps, to the velocity and direction of her throw. Throughout a range in which the velocity is still great enough for her rock to beat Billy's to the bottle, and Suzy's aim is still accurate enough to score a fairly direct hit, alterations to the throw will produce a counterfactually covarying range of alterations to the shattering. So the shattering is influenced by Suzy's throw.

Not so for Billy's throw. Unless we imagine Billy's throw sufficiently altered in time or manner so that his rock reaches the target before Suzy's does, the extent to which the time and manner of the shattering depends on alterations to the time and manner of the throw is completely negligible. This, claims Lewis, is why Suzy's throw counts as a cause of the shattering and Billy's does not.

Lewis also suggests that the analysis of causation as influence provides a solution to the problem of Trumping Preemption. Although there is no influence of the whether-whether or when-when variety, the transformation of the prince is nevertheless influenced by Merlin's spell since the manner of transformation covaries with the kind of spell cast. For example: had Merlin uttered "Presto! Prince-to-possum!" instead of "Presto! Prince-to-frog!" at noon, then the prince would have turned into a possum, rather than a frog, at midnight. On the other hand, what happened at midnight was in no way dependent on whether, when, or how Morgana acted in the late afternoon.

However this solution to the Trumping problem seems to turn on inessential and eliminable features of Schaffer's original example. If we suppose that Merlin's options are limited to a single spell that he may cast (standard prince-to-frog) and a single time of day he may cast it (noon), then the transfiguration of the prince is no longer influenced, in Lewis's sense, by Merlin's spell, though of course Merlin's spell is still its cause. (See Collins [2000], in Collins, Hall, and Paul [2004a, 114]. The idea is due to Jacob Rosen). This is a counterexample to the necessity of the influence theory. Other counterexamples to the necessity of the influence theory can be constructed from cases of Early or Late Preemption by ensuring that the preempted backup would have brought about the effect in exactly the same manner and at exactly the same time. See Strevens [2003] and Yablo's "Smart Rock" example reported in Hall [2004a, 237].

Schaffer has argued that the idea behind our first counterexample to necessity can also be developed into a counterexample to sufficiency. Add to the story of Merlin's limited options the fact that Morgana has a vast range of possible spells to cast, and available times at which to cast them. And now suppose Morgana stands by, silently watching, just before noon as Merlin prepares to cast his spell. Then, claims Schaffer, the transfiguration at midnight is influenced by Morgana's silent watching (given her vast range of options) though her watching is not among its causes. The point here seems less clear than in the previous case, since Schaffer must counter the suggestion that Morgana's watching *is* a cause — by *omission* — of the prince's becoming a frog. For more details, see Schaffer [2001]. Collins [2000] suggests that Lewis's own "Poison and Pudding" example was already a counterexample to the sufficiency of the causation as influence account.

5 DE FACTO DEPENDENCE.

In our Early Preemption example, the shattering of the bottle was not counterfactually dependent on Suzy's throw, which, nevertheless, caused it. Yet note this: holding fixed the fact that Billy does not throw his rock, the shattering *does* depend on Suzy's throw. That is, if Suzy hadn't thrown her rock and Billy had still not thrown his rock either, then the bottle would not have shattered. Stephen Yablo [2002; 2004] suggests we say that the shattering has a "de facto dependence" on Suzy's throw, and count Suzy's throw as a cause of the shattering in virtue of this de facto dependence. A version of Yablo's proposal will be developed in this section. Closely related accounts have been proposed by Judea Pearl [2000] and Christopher Hitchcock [2001].

In Late Preemption the same idea seems to work. Holding fixed the fact that Billy's rock doesn't hit the bottle, the shattering depends on Suzy's throw. That is, if Suzy hadn't thrown her rock, and yet Billy's rock still hadn't hit the bottle, then the bottle would not have shattered. Admittedly, this last counterfactual is a little weird. Since no backtracking is allowed, and we are holding fixed the fact that Billy's rock doesn't hit the bottle, the antecedent is asking us to suppose that Billy's rock is thrown just as it actually was, that it follows the same deadly-accurate trajectory toward the bottle (which is still there when it arrives, since Suzy hasn't thrown her rock) but then, somehow — miraculously! — fails to hit it.

Clearly some restrictions will have to be placed on the kind of proposition that may be held fixed, otherwise everything will turn out to depend on everything else given some suitably cooked-up background condition. For consider any simple and unproblematic case of causation: Billy acting alone throws a rock that shatters a bottle. Let *B* be the proposition that Billy throws the rock. Now let *L* be any other true proposition. *L* may be completely unrelated to, and independent of, the process that leads to the shattering of the bottle, for example the proposition that far away from the scene of the bottle breaking a certain leaf flutters down to the ground from a tree. Now let *G* be the proposition: *B* or not-*L*. Holding this proposition *G* fixed, the shattering of the bottle depends on the fall of the leaf.

In fact, for any pair of actual events whatsoever, we can simply let *G* be the proposition that either both of the events occur or neither does. Holding this proposition *G* fixed, each of the events depends on the other. Allowing the choice of these kinds of background conditions would trivialize the de facto dependence proposal. Yablo proposes to bar such choices on the grounds that are not sufficiently "natural", but he does not fully explain what naturalness amounts to in this context. The two examples just given suggest that it may be the disjunctiveness of the proposition in question that renders it unnatural, and that such background conditions are to be barred for the same reasons that disjunctive events are disallowed on many theories of events.

Even with this restriction, however, the proposal as sketched so far founders on cases of the following kind (due originally to Hartry Field):

(E4) **Self-Countering Threat** Billy places a bomb under Suzy's desk. Suzy notices the bomb and takes cover before it explodes. Suzy survives uninjured.

Holding fixed the true, and perfectly natural, proposition that there was an explosion under Suzy's desk, Suzy's survival de facto depends on Billy's planting of the bomb. That is, if Billy hadn't planted the bomb, and yet there had still been an explosion under her desk, then the bomb would not have been there for Suzy to notice, so she would not have taken cover, and hence would not have survived uninjured. Again, one should not be put off by the fact that this last counterfactual shares the strangeness of the corresponding counterfactual in our discussion of the Late Preemption example. Since no backtracking is allowed, and we are holding fixed the fact that there is an explosion under Suzy's desk, the antecedent is asking us to suppose that although there is nothing under Suzy's desk to lead her to take cover, there is, nevertheless, an explosion there.

Thus the proposal developed so far rules Billy's planting of the bomb to be a cause of Suzy's survival. But this seems quite incorrect, for it was Billy's action that threatened Suzy's survival in the first place. Billy's action can be thought of as a *self-countering threat* because it both threatens to prevent something, and also prevents that threat from succeeding. Intuitively: self-countering threats to an event's occurrence don't count as causes of it.

Yablo's strategy in the face of this problem is to attempt a principled explanation of the sense in which the dependence in these cases is "artificial".

His proposal is this: Suppose that E depends on C given H. Then say that the dependence of E on C given H is *artificial* if and only if every proposition that E depends on given that *not-C* is also something that E depends on given H. As Yablo puts it, the idea here is that the dependence on C is artificial just in case C "addresses some need" that is simply added to what would, if C had been false, have been all the needs.

Let's see how this works in our example of Self-Countering Threat. We want the dependence of Suzy's survival on the bomb planting, given the explosion, to turn out to be artificial. Is it? Well consider all those things that Suzy's survival would have depended on if Billy hadn't placed the bomb under her desk. Remember: we are now *not* holding fixed the fact of the explosion. In this "fallback scenario" as Yablo calls it, Suzy's survival depends on the usual mundane sorts of thing that all of us require in order to get through the day intact. It depends, for example, on her heart and lungs continuing to function normally, on her not being struck by lightning, and so on and so forth. Now the point is that all of these quite mundane things on which Suzy's survival depends in the fallback scenario, are also things on which her survival depends given that there is an explosion under her desk. No use avoiding the bomb blast and then suffering a fatal heart attack. It follows that the dependence of Suzy's survival on the planting of the bomb (holding the explosion under the desk fixed) is an artificial dependence.

But of course this fix is useless if it robs us of what we thought we had already obtained: a solution to the problem of Late Preemption. Returning to that example we must check that the dependence of the shattering on Suzy's throw, given that Billy's rock does not hit the bottle, does not also turn out to be artificial.

This dependence will prove to be artificial only if every event on which the shattering would have depended had Suzy not thrown her rock, is also an event that the shattering depends on given that Billy's rock doesn't hit the bottle. Now suppose that Suzy had not thrown her rock. Then the bottle would still have been shattered by Billy's rock, and the shattering would have depended on Billy's throw. But is Billy's throw also an event on which the shattering depends, given that Billy's rock does not hit the bottle? An odd question, perhaps, but one that is not difficult to answer in the negative if we keep the meaning of the "given that" locution firmly in mind. For clearly, if Billy had not thrown his rock, and Billy's rock had not hit the bottle, then the bottle would still have been shattered (by Suzy's) rock. So the de facto dependence of the shattering on Suzy's throw, given that Billy's rock does not hit the bottle, is not an artificial one.

Now anything can be made to look dependent on anything else, if we are allowed to hold fixed background conditions that are not sufficiently natural. That a similar trick might be used to show that every dependence is artificial, is the thought behind the framing of clause (D3) in the following formulation of Yablo's de facto dependence account of causation:

(D1) Say that *E depends on C given G* when the following counterfactual is true: If C were false and yet G were still true, then E would be false.

(D2) Say that the dependence of E on C given G is *artificial* when every proposition on which E would depend if C were false is also a proposition on which E depends given G.

(D3) Say that the proposition *E de facto depends* on *C* if and only if there is some true proposition *G*, such that *E* depends non-artificially on *C* given *G*, and such that G is more natural than any proposition *H*, given which the dependence of *E* on *C* is artificial.

Finally the proposal is that:

(D4) One event *c* is a *cause* of another event *e* if and only if the proposition that *e* occurs de facto depends on the proposition that *c* occurs.

One advantage that Yablo claims, if very tentatively, for the de facto dependence account, is that, unlike other approaches, it is "not at an absolute loss" when it comes to dealing with the Trumping Preemption examples. [Yablo, 2004, 134].

The idea here is that we can demonstrate the de facto dependence of the prince's transformation on Merlin's spell by holding fixed the proposition that no-one casts a spell before Merlin does. If Merlin had cast no spell then, given that no-one casts a spell before Merlin does, the prince wouldn't have turned into a frog.

Is this dependence artificial? Only if every event on which the transformation would have depended had Merlin cast no spell, is also an event the transformation depends on given that no-one casts a spell before Merlin. Is this true? Well, if Merlin had cast no spell, the prince's transformation would have depended on Morgana's six o'clock spell. But given that no-one casts a spell before Merlin, does the transformation depend on Morgana? No, because if Morgana had cast no spell, then Merlin would still have cast his (this being compatible with the proposition being held fixed) and the prince would still have turned into a frog.

But the real problem with this proposal for dealing with Trumping Preemption on the de facto dependence account seems to be with the degree of naturalness of the proposed background condition, rather than with the artificiality of the resulting dependence. To say that *no-one casts a spell before Merlin* is simply equivalent to saying that *either Merlin casts the first spell, or there is no spell cast at all.* The proposed condition is exposed as disjunctive!

But de facto dependence faces its own difficulties as well. It is not clear, for one thing, that Yablo has really succeeded in making clear the "*E* depends on *C* given *G*" locution; recall those weird counterfactuals that arose in the discussion of the examples. The same difficulty, it should be noted, also faces Pearl's and Hitchcock's version of the theory, though in those cases the problem is made less visible, since a solution to it is simply assumed to have been built into the framework in which a particular causal situation is to be modeled.

A second problem facing the theory has to do with the crucial notion of "naturalness", which Yablo has not fully succeeded in explaining.

Finally, Hall and Paul [forthcoming] have constructed a new and interesting case of preemption that appears to make trouble for the de facto dependence theory.

(E5) **Partial Preemption** Billy and Suzy are together attempting to push a heavy box across the floor. Billy is stronger than Suzy. He could shift the box all by himself if he really wanted to. Suzy couldn't. But Billy is also lazier than Suzy. When he sees that Suzy is using all of her strength, Billy decides to exert less effort than he might, and he applies a force insufficient on its own to move the box. Nevertheless their combined efforts succeed in shifting it.

Suzy's push and Billy's push are, intuitively, joint causes of the shifting. Yet the shifting fails to depend on Suzy's push, since had Suzy not pushed, Billy would have pushed harder and moved the box by himself.

Now let *G* be the proposition that Billy believes that he can exert less effort than he might, since Suzy is using all of her strength. Given *G*, the shifting depends on Suzy's push. But Yablo's account must deem this dependence artificial, since every event on which the shifting would have depended had Suzy not pushed, is also an event on which the shifting depends given that Billy believes he can exert less effort than he might.

6 THE BLUEPRINT STRATEGY

A third recent proposal that shows some promise in addressing the problems posed by the various preemption cases is the "blueprint strategy" developed and defended by Ned Hall (in [Hall, 2004a; 2004b]).

The idea behind the blueprint strategy is that we should hold fast to the thought that causation is an intrinsic matter; that whether or not a process is a causal one ought to depend only on the intrinsic nature of the process itself, and on the laws of nature.

In some ways the blueprint strategy might be thought of as a more sophisticated descendent of Lewis's 1986 account of causation as "quasi-dependence" (see [Lewis, 1986b, 193-212]) although, as we shall see, development of the strategy will turn out to require a quite radical break with the whole tradition of thinking of causation in terms of counterfactual dependence.

Here are the details of the quasi-dependence theory. Call a possible process that exactly matches all of the intrinsic properties of another process a *duplicate* of that process. If a possible process that is a duplicate of some process *P* exists in a possible world that has the same laws of nature as the world in which *P* is located, call it an *isonomic duplicate* of *P*. If causation is an intrinsic matter, then any isonomic duplicate of a causal process must also be a causal process. Suppose that some actual process is an isonomic duplicate of a possible process whose last event depends counterfactually on its first. Then say that the last event of the actual process is *quasi-dependent* on its first event. A *causal chain* is then taken to be any chain of events in which each event is either counterfactually dependent or quasi-dependent on the preceding event. One event is a *cause* of another if and only if there is a causal chain leading from the one to the other.

If this idea can be made to work, then it offers a ready solution to the problem of Late Preemption. The shattering of the bottle does not depend counterfactually on Suzy's throw, because of the backup process initiated by Billy when threw his rock a moment later. Yet it is not hard to find an isonomic duplicate of the process leading from Suzy's throw to the shattering, for which the counterpart shattering depends counterfactually on the counterpart of Suzy's throw. Just consider a possible scenario that is exactly like the actual one except that Billy does not throw his rock. The quasi-dependence theory thus correctly counts our case of Late Preemption as causal. Hall seeks to procure the same kind of solution to the preemption problem. In the example just discussed, the process initiated by Suzy's counterpart that leads to the shattering of the bottle's counterpart is an example of what Hall calls a *causal blueprint* for the actual process it duplicates. The blueprint process is uncontroversially causal, and we then appeal to the thesis that causation is an intrinsic matter in order to establish that the problematic actual process is a causal one too.

It is important here that a *process* be understood as something more than just an arbitrary chain or sequence of events. For it is simply not true that any isonomic duplicate of a chain of counterfactual dependence is a causal chain, as can be seen from the following example:

Dominoes: A large number of dominoes, standing on end, are arranged in a line so that when I push the first domino over it strikes the second, which falls and topples the third, and so on. The fall of the last domino is counterfactually dependent on the fall of the first one. But now consider a similar setup in which all except the first two and the last two dominoes have been removed. I topple the first domino so that it hits the second, and then later push over the second last domino so that it hits the final one. We can imagine this done in such a way that the pair of events consisting of the fall of the second and the fall of the final domino is an isonomic duplicate of the corresponding pair of events in the original setup. Yet the first of these two events does not cause the second.

Examining this case, we can that it is crucial that the structure of events duplicated include all the causal intermediaries of any pair of events in the structure. Less obvious, perhaps, is the fact that, unless *all* the causes contributing to the effect are included, a causal process leading up to some event may have an isonomic duplicate that is not a cause of the counterpart event, as the following example shows:

Voting: Billy, Suzy, and their friend Biff are voting on a proposal. Their votes are submitted electronically and a light goes on as soon as the machine receives either two yes votes or two no votes. Scenario (A): Billy and Suzy simultaneously press their yes buttons and the light goes on. Biff abstains. Scenario (B): Billy, Suzy, and Biff all press voting buttons simultaneously. Suzy votes yes, Billy and Biff vote no. The light goes on. Suzy's pressing her yes button was a (joint) cause of the light's going on in (A) but not in (B). (Based on a schematic example in [Hall, 2004b, 272–273].)

If we assume that there is neither backward causation nor causation at a temporal distance, this further condition may framed as the requirement that the structure of events to be duplicated consist of some event *e* along with *all* of the causes of *e* traced back to some earlier time.

One further adjustment needs to be made. The demand that the causal structures be perfect duplicates of one another is too strict. In our Late Preemption case, for example, the suggested blueprint process seems adequate despite the fact that it is a less than perfect intrinsic match of the original. But we are entitled to ignore the minute differences in the trajectory of Suzy's rock, due to subtle differences in the surrounding air currents, gravitational field and so on, that result from Billy's not throwing.

Putting all of this together, we arrive at Hall's formulation of the Intrinsicness thesis:

Intrinsicness: Suppose that S is a structure of events consisting of event e, together with all of its causes back to some earlier time. Let S' be a structure of events that intrinsically matches S in relevant respects, and that exists in a world with the same laws. Let e' be the event in S' that corresponds to e in S. Let c be some event in S distinct from e, and let c' be the event in S' that corresponds to c. Then c' is a cause of e'. [Hall, 2004b, 264]

Of course, this is only part of what has to be done. We want to claim that in problem cases like the preemption examples, causation amounts to duplication of a possible structure whose causal credentials are unproblematic (since the preempted backup processes are absent). But to complete the story, a defender of the blueprint strategy must explain what it is that constitutes the causal relation in the unproblematic situation in the first place. An obvious idea would be to appeal here to counterfactual dependence as a sufficient condition for causation in the blueprint, as Lewis hoped to do during the period in which he defended the quasi-dependence account. Then we could apply the Dependence thesis to establish that some blueprint is indeed a causal structure, and use the Intrinsicness thesis to carry this result over to solve the problematic preemption cases.

But unfortunately the obvious idea doesn't work, for the simple reason that the Intrinsicness thesis is incompatible with the claim that counterfactual dependence suffices for causation. Taken together they lead to unacceptable results. This can be seen from examples of causation by double prevention, cases in which one event causes another by preventing something that would have prevented the effect from occurring. And indeed the problem of double prevention was one of the reasons that Lewis abandoned the quasi-dependence account. (See [Lewis, 2004, 83–85].)

(E6) **Double Prevention**: A ball is thrown towards a fragile windowpane. Billy attempts to catch the ball to prevent it from hitting the window, but Suzy trips Billy, preventing him from catching the ball. The windowpane is shattered.

If Suzy hadn't tripped Billy, the windowpane wouldn't have been shattered. So if counterfactual dependence suffices for causation, the tripping counts as one of the causes of the shattering. But note that the causal credentials of the tripping rely partly on factors extrinsic to the actual structure S of events that includes the shattering and all of its causes traced back to the time at which the ball was thrown. For example, the tripping counts as a cause only because of the absence of anything *else* that would have prevented Billy from catching the ball, had Suzy failed to trip him. Suppose Biff had been present and ready to prevent the catch by shoving Billy aside. Then the tripping would *not* have been a cause of the shattering. Yet this new scenario (with Biff present) contains a duplicate of S embedded within it. Hence if we maintain both the Dependence thesis and the Intrinsicness thesis, we will misclassify the new scenario as also being one in which the tripping causes the shattering.

It would appear that either Intrinsicness or Dependence has to go. But Hall suggests a third way out. We may avoid the dilemma if we allow that the concept of cause is not univocal. He argues that there are in fact two concepts of causation [Hall, 2004a]. One of these — *causation as production* — satisfies Intrinsicness and is the concept of causation that the blueprint strategy is intended to elucidate. When we are inclined to judge that the causal relation is transitive, it is causation as production that we have in mind. On the other hand we have *causation as dependence*. When we have the intuition that double prevention is a kind of causation, when we are inclined to think that counterfactual dependence suffices for causation, and when we judge that an absence (the failure of events to occur) can both cause and be caused, it is causation in the sense of mere dependence that we are thinking of.

The blueprint strategy, then, applies only to causation in the sense of production. And it will involve a radical break from the counterfactual tradition. Hall currently suggests that we might use a version of the regularity theory, rather than counterfactual dependence, as a criterion for establishing the causal credentials of a blueprint. See Hall [2004b].

Are there really two distinct concepts of cause? One should, in general, be deeply suspicious of this sort of conceptual bifurcation strategy as a way of solving philosophical problems. As Lewis comments:

> the many concept hypothesis ...requires distinctions in our thinking that sometimes we do not make, need not make, and are in no position to make. If one event directly causes another, for instance, that is causation in one sense; whereas if one event causes another indirectly, in a case of double prevention (or in some still more indirect case) that is causation in a different sense. But when we neither know nor care whether the causation we have in mind is direct or indirect, what concept of causation are we employing then? [Lewis, 2004b, 286]

Lewis goes on to provide an example where we "neither know nor care" whether we are dealing with production or mere dependence. Air brakes on a train work by double prevention, vacuum brakes by production. Yet the judgment that one may cause the train to stop by pulling the emergency brake cord does not wait upon investigation of the contents of the black box to which the cord is attached. Jonathan Schaffer points out that the trigger mechanism of a gun also typically works by double prevention. No-one, however, is tempted to conclude from this that pulling the trigger of a gun cannot cause it to fire, or only causes it to fire in some secondary sense of "cause". (See [Schaffer, 2000b; Maudlin, 2004, 438].)

BIBLIOGRAPHY

[Albert and Loewer, 2005] D. Albert and B. Loewer. Counterfactuals and the second law, in Huw Price and Richard Corry (eds.) *Russell's Republic: the Place of Causation in the Constitution of Reality*, Oxford: Oxford University Press, 2005.

[Collins, 2000] J. Collins. Preemptive prevention, *Journal of Philosophy* 97: 223-234, 2000. Reprinted in Collins, Hall and Paul (eds.) 2004a: 107-117.

[Collins *et al.*, 2004a] J. Collins, N. Hall, and L. A. Paul (eds.). *Causation and Counterfactuals*, Cambridge, Mass.: MIT Press, 2004.

[Collins *et al.*, 2004b] J. Collins, N. Hall, and L. A. Paul. Counterfactuals and causation: history, problems, and prospects, in Collins, Hall, and Paul (eds.), 1-57, 2004.

[Elga, 2000] A. Elga. Statistical mechanics and the asymmetry of counterfactual dependence, *Philosophy of Science* (suppl. vol. 68, PSA 200): 313-324, 2000.

[Goodman, 1947] N. Goodman. The problem of counterfactual conditionals, *Journal of Philosophy* 44, 113-128, 1947. Reprinted in Nelson Goodman *Fact, Fiction, and Forecast*. Cambridge, Mass.: Harvard University Press (1979).

[Hall, 2000] N. Hall. Causation and the price of transitivity, *Journal of Philosophy* 97: 198-222, 2000. Reprinted in Collins, Hall, and Paul (eds.) 2004a: 181-203.

[Hall, 2004a] N. Hall. Two concepts of causation, in Collins, Hall, and Paul (eds.), 225-276, 2004.

[Hall, 2004b] N. Hall. The intrinsic character of causation, in Dean Zimmerman (ed.) *Oxford Studies in Metaphysics*, Vol. 1, Oxford: Oxford University Press. 255-300, 2004.

[Hall and Paul, forthcoming] N. Hall and L. A. Paul. *Causation and the Counterexamples: A Traveler's Guide*, Oxford: Oxford University Press, forthcoming.

[Hitchcock, 2001] C. Hitchcock. The intransitivity of causation revealed in equations and graphs, *Journal of Philosophy* 98, 273-299, 2001,

[Hume, 1902] D. Hume. *An Enquiry Concerning Human Understanding*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press, 1902 (1748)..

[Lewis, 1973a] D. Lewis. Causation, *Journal of Philosophy* 70: 556-67, 1973. Reprinted in Lewis 1986a: 159-172.

[Lewis, 1973b] D. Lewis. *Counterfactuals*, Cambridge: Harvard University Press, 1973.

[Lewis, 1979] D. Lewis. Counterfactual dependence and time's arrow". *Noûs* 13: 455-476, 1979. Reprinted with Postscripts in Lewis 1986a: 32-66.

[Lewis, 1986a] D. Lewis. *Philosophical Papers, Volume II*, Oxford: Oxford University Press, 1986.

[Lewis, 1986b] D. Lewis. Postscripts to 'Causation', in [Lewis, 1986a, 172-213].

[Lewis, 1986c] D. Lewis. Causal explanation, in [Lewis, 1986a, 214-240].

[Lewis, 2000] D. Lewis. Causation as influence (abridged version), *Journal of Philosophy* 97, 182-197, 2000.

[Lewis, 2004a] D. Lewis. Causation as influence (full version), in [Collins *et al.*, 2004a, 75-106].

[Lewis, 2004b] D. Lewis. Void and object, in [Collins *et al.*, 2004a, 277-290].

[Mackie, 1965] J. L. Mackie. Causes and Conditions, *American Philosophical Quarterly* 2, 245-264, 1965.

[Maudlin, 2004] T. Maudlin. Causation, counterfactuals, and the third factor, in [Collins *et al.* 2004a, 419-443].

[McDermott, 1995] M. McDermott. Redundant causation, *British Journal for the Philosophy of Science* 46, 523-544, 1995.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press, 2000.

[Schaffer, 2000a] J. Schaffer. Trumping preemption, *Journal of Philosophy*, 97: 165-181, 2000. Reprinted in [Collins *et al.*, 2004a, 59-73].

[Schaffer, 2000b] J. Schaffer. Causation by disconnection", *Philosophy of Science* 67, 285-300, 2000.

[Schaffer, 2001] J. Schaffer. Causation, influence, and effluence, *Analysis* 61: 11-19, 2001.

[Stalnaker, 1968] R. Stalnaker. A theory of conditionals. In Nicholas Rescher (ed.), *Studies in Logical Theory*. Oxford: Blackwell, 1968.

[Strevens, 2003] M. Strevens. Against Lewis's new theory of causation: a story with three morals, *Pacific Philosophical Quarterly* 84: 398-412, 2003.

[Yablo, 2002] S. Yablo. De facto dependence, *Journal of Philosophy* 99: 130-148, 2002.

[Yablo, 2004] S. Yablo. Advertisement for a sketch of an outline of a prototheory of causation, in [Collins *et al.*, 2004a, 119-137].

LOGIC, MATHEMATICS, AND THE NATURAL SCIENCES

Neil Tennant

1 INTRODUCTION

This paper is in two parts. The first part sets out a 'minimalist' position regarding the correct logic by means of which one may pursue the hypothetico-deductivist method in natural science. It is argued that *intuitionistic relevant logic (IR)* is adequate. A requirement of reflexive stability is that such an argument should be conducted within the confines of the very logic whose methodological adequacy is to be established. The second part of the paper (§6) addresses an objection in principle that has been raised by John Burgess [2005a], [2005b]) to the reflexive stability of the argument for *IR*.

2 THE QUESTION OF THE ADEQUACY OF CONSTRUCTIVE METHODS FOR HYPOTHETICO-DEDUCTIVISM IN NATURAL SCIENCE

Dummett [1991, 320–1] writes

> Investigation might reveal that a constructivist version of a mathematical theory was perfectly adequate for the applications made of it within natural science... These questions have scarcely been raised, let alone answered, by either mathematicians, philosophers or physicists.

This claim of Dummett's is not quite accurate. Investigation had already revealed, at least by the time he published the claim, that a constructivist version of a mathematical theory is adequate for all the applications to be made of the theory within natural science.[1] Or, to be more precise: investigation had already revealed that, in the overall context of the hypothetico-deductive method in natural science, constructivist *logical reasoning* is adequate for the applications to be made even of *classical* mathematics. Indeed, the investigation reveals exactly how and why *one does not err*, even by constructivist lights, in deriving and applying strictly

---

[1] See [Tennant, 1987], Chapter 18, 'Intuitionistic Relevant Logic is Adequate for Popperian Science'. Those ideas were developed further in [Tennant, 1997], Chapters 12 and 13. The original source of publication of the treatment in Chapter 18 of [Tennant, 1987] is [Tennant, 1985].

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

classical mathematical theorems; and how and why, for all the use that might be made of strictly classical mathematics, it is a ladder that the constructivist can kick away and ultimately forswear.

This claim is complicated, and perhaps surprising. The aim here is to make it clearer and completely plausible.

The basic idea is that, in so far as one is concerned to be able to produce all possible refutations of empirical theories, the underlying logic can, without loss of generality, be taken to be very weak indeed. Either Johannson's system of minimal logic or the author's system of intuitionistic relevant logic will do. The mathematical theorems that are applied in the course of physical theorizing turn out, on this account, to be deductive halfway houses: they stand as conclusions of purely mathematical arguments (often strictly classical ones), but stand as premises in the applied scientific context within which an empirical refutation is formulated. Upon constructivizing (and relevantizing) the *reductio* involved in the refutation of an empirical theory, the erstwhile mathematical theorem might be elided. Yet the mathematical axioms will still be in force, helping to sustain the overall reductio. And the logic of the reductio, is, to repeat, weak.

This provides a satisfactory affirmative answer to the question whether 'a constructivist version of a mathematical theory [i]s perfectly adequate for the applications made of it within natural science'. It also affords a constructive explanation of why it is that one can trust strictly classical mathematics when it is applied in the context of empirical theorizing. All that the classical aspect of the underlying logic is securing is a possibly shorter, more convenient proof. But, as the overall result shows, empirical refutations rely only on the constructive content of the underlying mathematical axioms. Strictly classical mathematical results that are applied within scientific theorizing are — to repeat — ladders that can be kicked away, once they have served their expeditious purpose (in securing the predictions that are eventually refuted).

3 REFUTATION OF EMPIRICAL THEORIES

When an explicitly formulated theory fails a well-defined and well-controlled experimental test, it is refuted. Moreover, *so long as we trust the evidence and the controls that applied to the experimental conditions*, it stays refuted. Note that by saying that it is the theory (that is, the collection of high-level hypotheses) that has failed the test, we are already supposing that the Poincaré–Duhem–Quine problem has been resolved. That is, we are agreed that the boundary condition statements were true (the experiment was properly controlled) and that the experimental observations were correct (we made no mistakes in reading our measuring apparatus, and it functioned correctly, etc.). Since what was observed conflicted with what was predicted (by flawless logical reasoning) from the theory and the boundary condition statements, we have concluded that it was the theory that was at fault:

$$\begin{array}{c}
\underbrace{\begin{array}{l}\text{Theory (hypotheses) +}\\ \text{Boundary condition statements +}\\ \text{Auxiliary assumptions about measuring apparatus}\end{array}}\\
\vdots\\
\underbrace{\text{Predictions , Observations}}\\
\vdots\\
\perp
\end{array}$$

If we do not revise our evaluations of past evidential statements in the light of any new ones, then future consistent extensions of present evidence will leave the refuted theory refuted.

The slogan 'Once proved, always proved' in mathematics can be explicitly restricted (without loss) so as to encompass only *consistent* extensions of our present axiomatic basis. That is to say, there is by default a consistency requirement on the axiomatic basis of mathematics. In the logico-mathematical case: if axioms X enable us to prove theorem A, then future axioms $X{+}X'$ — which of course we assume to be consistent — enable us to prove theorem A.

There is also by default a similar consistency requirement on whatever is supposed to constitute the 'total evidence' for or against any empirical theory. In every domain, consistency is a prerequisite for truth.

The slogan 'Once empirically refuted, always empirically refuted', can be justified analogously as follows: if evidence E refutes theory T, then future evidence $(E{+}E')$ — which of course we assume to be consistent — also refutes theory T.

With a logico-mathematical assertion P there can be neither retraction nor revision, unless some mistake is revealed in what had been taken as a proof of P. 'Once proved, always proved' is the motto here. Knowledge (properly pedigreed) is strictly cumulative over time. And the deductive logic governing such assertions is *monotonic on its premises*. That is, if X is a subset of Y, and X logically implies A, then Y logically implies A. In the logico-mathematical case, the subset X of premises that consistently expands to become the set Y may be thought of as the foundational basis for one's assertions. Such sets provide the ultimate grounding or support for one's conclusions A. When X consists of mathematical axioms only, we say that A is a mathematical theorem. (When X is empty, we say that A is a logical theorem.) Our certainty in the mathematical axioms transmits, via logical deduction, to the theorems as well. Certainty transmits along lines of logical implication, from premises to conclusions.

The deductive logic governing empirical claims based on empirical evidence is *also* monotonic, *despite* the fact that the history of scientific theorizing may reveal 'oscillations' in the fortunes of either an empirical hypothesis, or an evidential statement. The claim about monotonicity here is *not* to be interpreted as resting on a naïve epistemological analogy between mathematical axioms and evidential statements. *Of course* the former are far more certain than are the latter. Indeed, mathematical axioms are often *self*-evident and necessary; whereas evidential statements are never so. The crucial difference between the two cases, however, is this. In the empirical case, unlike the mathematical case, the direction of logical implication (with respect to which such monotonicity holds) runs *counter to* the direction of evidential support. When we 'base' a (consistent) scientific theory X on empirical evidence E (consisting of true observation statements) we require (roughly)[2] that X should logically imply E, not vice versa. This logical implication is, to be sure, monotonic on the premises. But the premises X (forming our scientific theory) are what, in this context, require the support! And it is the conclusion E that is supposed to provide it. The support that E affords X is not just a matter of X's logically implying E. The latter is necessary, but not sufficient, for the empirical evidence E to support the empirical theory X.

Another condition is that X should (perhaps in conjunction with further assumptions B) logically imply further observational conclusions E^*. The observation sentences E^* should be ones concerning which we have as yet no firm opinion. Such observational conclusions E^* logically inferred from X and B are the predictions of X (modulo B) that can be used to test X. The test will take place by arranging circumstances for which B holds, and then seeing whether E^* holds too. The conditions B are called *boundary conditions* for such experimental testing. When the experiment is conducted, there are, ideally, two outcomes: E^* is seen to hold, or E^* is seen not to hold. In the former case, the theory X is corroborated, for X's prediction (that E will hold in circumstances B) is borne out. Moreover, E will also have been *explained* by X in the circumstances B. In the latter case, where E is seen not to hold, the premises X are collectively refuted. We shall return presently to consider how decisive such refutation may be for any particular member of X. Just which members of X ought to be retracted in the light of such a refutation of X as a whole is in general a rather vexed question.

A third necessary condition for the empirical evidence E to support the empirical theory X is that the evidence within E should be gathered from various domains, and that the formulation of X should be economical and abstract enough not to reflect such diversity explicitly, but to do so only implicitly, by way of deductive application via the boundary conditions B. That is, X should provide *unifying explanations of disparate phenomena.* X should be of wide evidential scope. An example would be the way that Newtonian dynamics and the theory of gravitation (as such X) can be applied to provide explanations of such diverse phenomena as apples falling from trees, projectiles following roughly parabolic trajectories, spinning tops precessing, ocean tides correlating with the position of the Moon, the motions of pendula, the orbits of the planets, the efficacy of aerofoils, the vibrations of a plucked string, the formation of sand-dunes, ripples on a pond, sonic booms, and many other such phenomena.

---

[2] Quine would prefer to say that we require X to yield 'pegged observational conditionals' of the form 'if O_1 and ... and O_n then O'. Some of the evidence E has to be accepted at face value, in the form of various O_i. The rest of E would then have to be covered via detachment using these conditionals. Obviously there will be a premium on minimizing the former and maximizing the latter. *Cf.* [Quine, 1990].

A fourth necessary condition (which is related to the third) is that X should be *simple*. This is a notoriously difficult virtue to analyse or explicate, but scientists do have strong intuitions about whether particular theories are simple. Theories can fail to be simple in various ways: they can be too *ad hoc*; they can amount to little more than restatement of the evidence; they can postulate too many kinds of hidden entity to perform the explanatory job at hand; they can extrapolate from their data points in 'unsmooth' ways. We shall not be too concerned here to detain the reader with any attempt to explicate simplicity further. It is not important, for present purposes, that one be able to do so.

We were considering the logic of scientific explanation, and how the direction of theoretical explanation runs counter to the direction of evidential support. Theory X in circumstances B explains evidence E only if E provides evidential support, in circumstances B, for the theory X. And this amounts to no more than E following logically from the conjunction of X with B. We do not intend to say anything about how the evidence mounts up, or about how certain evidential statements can be more important than others. We do not offer the prospect of any further metatheoretical development of the relation of support, either in the form 'E_1 would be better evidence for X than would E_2' or in the form 'E would be better evidence for X_1 than it would be for X_2'. Again, this is not needed for present purposes. Nor do we intend to say anything about confirmation or probabilification of hypotheses by evidence, and Bayesian conditionalization.

It is enough to confine our treatment to the strictly hypothetico-deductive model of explanation. For it is clear that it is a workable model in so far as it goes. The extent to which it does not accommodate all the intuitions that scientists and methodologists may have about how theories relate to the evidence does not concern us. All that is important is that one recognize the fundamental features captured by the hypothetico-deductive model. It is quite conceivable that there should be thinkers and reasoners who exploited and relied on those features and those features alone. Their thought about the external world and its deep regularities, and their quest for theoretical explanations of empirical phenomena, could be just as cogent and urgent as ours. Moreover, their scientific successes could be just as impressive as ours, and they could use their theories just as we do, as guides to life, as a means to anticipate courses of events, as considerations in choice of future actions, and as sources of technological innovation.

We are *not* trying to provide a full account of human scientific rationality. That would be way beyond the scope of our concerns. Rather, we are isolating an essential logical core to our competence as empirical theorizers — a core which *could*, arguably, serve as the total competence of some species of rational agent, in so far as empirical theorizing is concerned, even if it falls short as an account of our own full competence in that regard.

Given this strictly limited and modest concern, we are accordingly under no obligation to essay upon the abductive 'logic' of *discovery* or of scientific invention. We offer no account of how a scientific intellect, confronted with a range of evidence, would come up with a high-level theory that successfully explains it and that can be tested against the further predictions that it makes. We are content to leave that process mysterious and untouched. We are interested only in what happens, logically, after the theory has been formulated. We are interested here only in the deductive logic of theory *testing*.

Similarly, we are content to deal only with a very regimented language, namely, the language of first-order logic. Even if this language should prove (*pace* Quine) to be inadequate for the expression of all our thought about the empirical world, it is nevertheless clear that there can be systematic thought, framed in a first-order language, about the empirical world, and that the essential features of the hypothetico-deductive model of explanation will be in place when the logical deducibility relation[3] in that language is taken as the relation involved in prediction and explanation.

Every logically contingent statement that ventures logically beyond the present observational evidence is vulnerable to empirical refutation. But usually it is vulnerable in the company of others. When we have derived absurdity from a set of evidential statements in conjunction with a set of conjectural statements, we have many ways of proceeding from there. We may retract one or more of the evidential statements; and likewise with the conjectural statements. In general, we might be able to contract to any one of several consistent subsets of (evidence+conjectures). How we respond to refutations (proofs of absurdity) is, again, a matter more of the 'logic' of scientific discovery than it is a matter of the logic of testing. The refutation consummates the test; something has failed. Whether the failure is to be located among the conjectures venturing beyond the evidence, or in the so-called 'evidence' itself, is a matter that need not detain us. Let Δ contain all the premises involved in the refutation. Thus Δ embraces conjectures (hypotheses) and evidential statements alike. The refutation of Δ is decisive, in the sense that any further evidence Γ leaves it untouched. If, now, we hold to the evidential statements in Δ, the hypotheses forming the rest of Δ stay jointly refuted, despite the accumulation of the new evidence Γ. So: as the tribunal of experience recruits more members (observation statements held true) and speaks with one voice (is logically consistent), we can say of the theories that founder on this evidence: once refuted, always refuted.

4 THE ANTI-REALIST CONSTRUAL OF EMPIRICAL CLAIMS THAT CANNOT BE PROVED

The anti-realist content of the unprovable empirical generalization that all F's are G's is really this: *Nature will not confound the assertion that all F's are G's.* That is, Nature will not yield a case of an F that is not a G. In general, the anti-realist content of any empirical assertion P of which, by virtue of its general or hypothetical nature, we can say *a priori* that it cannot admit of proof, is: *Nature will not refute* P.[4] For such a belief will only ever be entertained as an explanatory hypothesis, and as a generator of predictions; and, as such, is subject only to refutation, not proof. Refutation, however, would always be modulo some set of assumptions that were firmer than the belief P in question. The Poincaré–Duhem–Quine problem is simply that of how we focus on the particular P that we might take to have been refuted, once we have a disproof of a set of assumptions containing it. But that is a problem for applications. All that our deductive logic can be expected to provide is the various disproofs of these sets of assumptions as the intellectual opportunity or need arises. Now to this end, the construal of any (unprovable) empirical assertion P as being to the effect that Nature will not refute P makes the system *IR* of intuitionistic relevant logic perfectly adequate for empirical science, with its deductive testing of explanatory and predictive hypotheses against the evidence. If we agree that we may turn any claim of the form $\forall x(Fx \rightarrow Gx)$ into the corresponding form $\neg\exists x(Fx \wedge \neg Gx)$, then we can supply in *IR* all the disproofs needed for empirical science. For, in the language based on $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\exists$, we have the metatheorem:

---

[3] Or perhaps some suitably constrained subrelation thereof.

METATHEOREM 1. If Δ can be disproved in classical logic, then Δ can be disproved in *IR*.

That is all very well, says the objector; but what about the case where one is drawing out a logical consequence in the form of a prediction which has not yet been refuted? How does the anti-realist using only *IR* match that? The answer is that if one has derived the prediction P from the assumptions Δ using classical logic, then in *IR* one can at least derive $\neg\neg P$ from Δ. $\neg\neg P$ is the regimentation of 'Nature will not refute P'. And this is the appropriate propositional attitude to have, according to the anti-realist, towards the prediction P, which cannot, on the basis of the present evidence, admit of proof. The 'proof' by means of which we make the prediction P will of course involve as undischarged assumptions the higher-level hypotheses of our explanatory empirical theory; and since these assumptions cannot admit of proof, nothing that depends on them for 'proof' is really proved. The only genuine proof one could ever have for a prediction P would be based on various atomic axioms that will only be available in the future, once events have run their course. Our theories can enjoy no proof in the present, but at best withstand the test of time. If and when the countervailing evidence comes in the form of $\neg P$, however, then there is nothing to choose between the following two logical passages:

---

[4] The classicist is committed to this immediately by maintaining that 'All F's are G's' is logically equivalent (in particular, entailed by) 'It is not the case that some F is not a G'. Popper, for example, urges that one use the latter *in place of* the former when regimenting our scientific theories. For the anti-realist, for whom 'It is not the case that some F is not a G' does not in general entail 'All F's are G's', greater discrimination is called for. Thus when the latter form of words is used for the formulation of a scientific hypothesis, our proposed reading secures the logical licence that is generally withheld.

$$\dfrac{\neg P \qquad \begin{matrix}\Delta\\ \vdots\ \text{via classical logic}\\ P\end{matrix}}{\bot} \qquad\qquad \dfrac{\begin{matrix}\Delta\\ \text{via } IR\ \vdots\\ \neg\neg P\end{matrix} \qquad \neg P}{\bot}$$

Another objection worth disposing of here is the following allegation of circularity: 'You say that the content of an empirical assertion P of which, by virtue of its general or hypothetical nature, we can say *a priori* that it cannot admit of proof, is: *Nature will not refute* P. But to understand the latter, we need first to understand P itself.' This objection implies that our account is unable to provide the content P independently, so that it can later be embedded in the context 'Nature will not refute . . . '.

But this is to misunderstand the overall division of conceptual labour. The content of P is already available via composition out of the meanings of its constituent expressions. The latter, in turn, have had their meanings conferred on them by the inferential liaisons that they enjoy within the empirical theory (if they are empirical terms) or by the rules of inference that govern them (if they are logico-mathematical terms). That yields us the *sentential content* P. Now, when we advance to consider an *assertion of* P, we are free to append a further analysans to P, in order to capture the special illocutionary force involved.

5 THE FATE OF STRICTLY CLASSICAL THEOREMS OF APPLIED MATHEMATICS

Past debate over the adequacy of constructive or intuitionistic mathematics for natural science has focused on the question whether particular theorems finding application in natural science can be proved by strictly constructive or intuitionistic means. A standard way of alleging the inadequacy of constructive methods would be as follows:

> The classical mathematical theorem φ finds application within such-and-such branch of empirical science. (It is used to explain phenomena, derive predictions, etc.) But φ is not a theorem of constructive mathematics. Hence constructive mathematics is inadequate for the demands of actual, current, scientific practice.

The constructivist might challenge the claim of constructive unprovability (as, for example, [Bridges and Richman, 1999] *contra* [Hellman, 1993]). But even if that were to fail, another possible response by the constructivist might then be to show that some constructive theorem 'close enough' in spirit to φ might nevertheless suffice, given certain circumstances revealed by inspection of the mathematics in question. For example, it might be argued that it would suffice for predictive purposes if one could approximate the classicist's real values by means of intuitionistically ratifiable rational values that are accurate to within a reasonable margin.

The situation regarding constructivism is analogous to that regarding relevantism. Relevant logicians have held out the prospect of 'relevantizing' mathematics, that is, of deriving all mathematical theorems from the axioms by means of their favored system of (classical) relevant logic. The underlying contention is that intuitive mathematical reasoning is always fundamentally 'relevant', hence that its formalization in a suitably chosen relevant logic should not result in any significant loss of mathematical theorems. The problem, however, especially for proponents of systems of relevant logic lying in the neighborhood of the system R of Anderson and Belnap, is that they could relevantize theories such as arithmetic only piecemeal. They lacked any general metatheorem to the effect that every classically proved theorem could be proved from the axioms by means of relevant logic.

The difference, in the intuitionistic case, is that of course we know that certain theorems of classical mathematics will be lost upon adoption of an intuitionistic or constructive logical system. So the intuitionist who wishes to claim that there is no great loss as far as *applied* mathematics is concerned needs to take a closer look at the exact logical needs of the scientific theorist seeking to apply mathematical theorems in pursuit of empirical predictions that may be tested by observation and experiment.

The line of argument proposed in this paper is intended to accomplish two aims. First, it is intended to bypass or circumvent the sort of dispute between the classicist and the constructivist that has taken place between Hellman and Bridges. As Bridges [1999, 440] puts it,

> **constructive mathematics is none other than mathematics carried out with intuitionistic logic.**

And, one might add (by way of generalizing this dictum),

> **constructive hypothetico-deductive theory-testing is none other than hypothetico-deductive theory-testing carried out with intuitionistic logic.**

The second aim is to show in one fell swoop, rather than piecemeal, that there is absolutely no loss of intuitionistic (resp. classical) mathematical content upon adopting the relevant logic *IR* (resp. *CR*).

Let us take a closer look, then, at the logical needs of the scientific theorist. The aim of the hypothetico-deductivist is to derive, from the scientific hypotheses being tested, predictions that may conflict with, or disagree with, or contradict one's observations and measurements. The overall structure of such a *reductio ad absurdum* is as follows:

$$\begin{array}{c}
\text{Mathematical axioms} \\
\Pi \\
\underbrace{\text{Mathematical theorems, Scientific hypotheses, Boundary conditions}} \\
\Sigma \\
\underbrace{\text{Predictions, Observations/measurements}} \\
\Theta \\
\perp
\end{array}$$

A mathematical theorem might be needed, say, to provide a characteristic form of solution for the particular kind of differential equation that might be used in the statement of a scientific hypothesis. Such an hypothesis might deal with the time-evolution of various co-varying physical magnitudes, stating in mathematical terms the precise relationships among them, and how they change with respect to one another. The strictly mathematical proofs Π furnish the mathematical theorems (from one's mathematical axioms) that are desired for such applications within natural science. From the point of view of the practising scientist, unconcerned with foundational matters in mathematics, it is good to have a ready supply of mathematical theorems 'off the shelf', whose *a priori* credentials have been taken care of by the mathematicians. The theoretical scientist who simply 'applies' the available mathematics is usually concerned only with the next level down in our schema: with the subsequent derivations Σ of predictions (about the outcome of possible experiments), based on the scientific hypotheses and statements of whatever boundary and initial conditions will be controlled for in those experiments. It is then left to the experimentalists to carry out such experiments, and to gather and analyze the resulting data. Any eventual conflict between prediction and observation would be brought out at the lowest level in our schema, in the derivation Θ of absurdity ($\perp$). If the prediction concerned, say, features of the statistical distribution of measured values, then the data-processing involved in showing that the prediction is confuted might be quite complex; but Θ would still be a relatively low-level proof within computational mathematics. If, however, the prediction concerned some simple observable, such as 'It will land here', and the observation turned out to be 'It landed way over there', then Θ would consist of but a single step of inference to $\perp$.

Unlike the mathematician, who is concerned only with the subproof Π, and unlike the theoretical scientist, who is concerned only with the subproof Σ, and unlike the experimentalist, who is concerned only with the subproof Θ, the philosophical methodologist is concerned with the whole schema consisting of Π, Σ and Θ. A question that naturally arises is whether, whenever a *strictly classical* theorem φ

stands as the conclusion of some classical subproof Π, and as an assumption for some (possibly classical) subproof Σ:

$$
\begin{array}{l}
\text{Mathematical axioms}\\
\qquad\qquad\Pi\\
\underbrace{\text{Mathematical theorem } \varphi\text{, Scientific hypotheses, Boundary conditions}}\\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Sigma\\
\qquad\qquad\qquad\qquad\qquad\qquad\underbrace{\text{Predictions, Observations/measurements}}\\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Theta\\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\bot
\end{array}
$$

on the way to a derivation of ⊥ via the proof Θ that draws on the outcomes of observations, *there might not be an alternative, fully constructive, deductive route to ⊥ from the same set of overall assumptions.*

The answer to this question is affirmative, and fully general. Any *reductio ad absurdum* that proceeds via a 'strictly classical' mathematical theorem can be effected with intuitionistic relevant logic. That is, the system *IR* will furnish a proof of ⊥ from the combined premises: mathematical axioms, scientific hypotheses, initial and boundary conditions, and observations/measurements. In such an *IR*-proof, *the passage via the strictly classical theorem φ will have been avoided altogether.* The classical theorem is never, strictly speaking, necessary for the precipitation of ⊥! Indeed, it is never, strictly speaking, necessary for the derivation of any decidable prediction. Since predictions are always of decidable matters of fact, we can summarize by saying that *strictly classical mathematical theorems are in principle otiose for the purposes of natural science.*

This is a startling result; but it is guaranteed by our constructively provable Metatheorem 1.

Intuitively, what happens is this. One takes the fully formalized proof

$$\Pi(\varphi)\Sigma(\mathit{Prediction})\Theta\bot$$

and one constructivizes it, using the constructive method furnished by the constructive proof of the Gödel–Gentzen–Glivenko theorem. Call the resulting proof Ξ. (Note that Ξ will not contain any subproof establishing φ as a theorem, since, *ex hypothesi*, φ is a strictly classical theorem.) Then one normalizes Ξ (a process which in general takes superexponential time, as a function of the length of Ξ); and finally one extracts the relevant kernel of Ξ (i.e., a proof in the system *IR*) by applying the linear-time transformation described in [Tennant, 1992] (for the propositional case) and [Tennant, 1994] (for the first-order case). What this transformation does, essentially, is delete all applications of the Absurdity Rule (*Ex Falso Quodlibet*), and slightly massage applications of remaining rules (such as conditional proof, and proof by cases) so that they count as applications of those rules as stated for *IR*.

The constructivization and relevantization of *reductio* proofs described above is made possible by (i) the embeddability, via a suitable translation, of classical logic into intuitionistic logic, and (ii) the relevantizability of proofs with epistemic gain. The fully general form of (ii) is given by the following metatheorem.

METATHEOREM 2. If φ can be deduced from Δ in classical [resp., intuitinistic] logic, then either φ or $\perp$ can be deduced from (some subset of) Δ in the classical [resp., intuitionistic] system *IR*.

It is an important feature of proofs in the systems *IR* and *CR* that they are in normal form (if natural deductions) or cut-free (if sequent proofs). Moreover, the natural deductions in either system may not contain applications of the absurdity rule; and the sequent proofs in either system may not contain any dilutions (or weakenings). The two systems otherwise involve only slight tweaking of the rules for the logical operators, so as to offset the deductive sacrifices that would otherwise have to be made upon eschewing the types of inferential moves just mentioned. (For a fuller treatment of both the natural deduction and sequent formulations of *IR* and *CR* see Tennant [1997, ch. 10].

6 MEETING AN OBJECTION OF BURGESS

In [Burgess, 2005a, 734–740], Burgess sets out a summary of the foregoing position, which he calls *perfectionism*, and enters an important (but rhetorical) objection. (The gist of his criticism is reprised, *con brio*, in [Burgess, 2005b].) The label 'perfectionism' is perfectly welcome, if it helps to identify the position in its essentials.

Burgess rightly points out that deductive progress in mathematics is cumulative, involving the interpolation of lemmas *en route* to mathematical theorems from one's axioms. Typically we chain together proofs, choosing lemmas so as to reduce the deductive workload — massively. That way, theorems are put within reach that would otherwise be practically inaccessible, were we to limit ourselves to cut-free proofs not involving inferential passage to, and then from, our chosen lemmas. Those are the facts of mathematical life; that is how mathematicians actually proceed.

So, given that all the proofs in the relevant systems *IR* and *CR* are cut-free, it would not in general be practicable to insist that proofs of mathematical theorems should actually be carried out within *IR* or *CR*. Let us focus on classical mathematics and the system *CR*, for the present purpose of discussing Burgess's objection to the reformist's contention that *CR* is the correct (because suitably *relevant*) system of logic for classical mathematics. (Note that at this point the discussion focuses on mathematics; but we shall not lose sight of the broader theme of scientific reasoning in general, and whether *IR* is adequate for it.)

The reformist (or 'perfectionist', in Burgess's terminology) claims that there is no loss incurred by restricting oneself to CR — on the contrary, there is possible epistemic gain to be had. (See Metatheorem 2 above.) Burgess maintains that this claim is 'unsurprisingly not uncontroversial' (p. 739, fn. 6). His evaluation of the dialectical situation is as follows (pp. 739–740):

> ... in general, Cut can be eliminated only at the cost of making proofs infeasibly long.[fn] So even though *in principle* almost anything classically provable will be perfectionistically provable or else something even better will be, *in practice* the proofs mathematicians actually give not only *do* fail to adhere to the restriction of perfectionsim, but *must* do so if they are to be kept to a humanly comprehensible length.
>
> The perfectionist might then reply that in practice mathematicians could go on working as they do now, since Tennant's work shows that in principle everything or almost everything they are doing could be justified from a perfectionist standpoint. The trouble with this response is that it relies on a theorem ... for which only a classically and not a perfectionistically acceptable "proof" has been given. How far can a logician who *professes* to hold that perfectionism is the correct criterion of valid argument, but who freely accepts and offers standard mathematical proofs, in particular for theorems about perfectionist logic itself, be regarded as *sincere* or *serious* in objecting to classical logic? This question will be left open for the reader to ponder ... .

Now, all that is relevant to an assessment of the disagreement between Burgess and myself is the following set of high-level assertions:

1. The reformist (relevantist) R is claiming that *the correct logic* is some system $\mathcal{R}$ (or set $\mathcal{R}$ of proofs) properly contained in the system $\mathcal{C}$ of the conservative (classicist) C.
2. R, the reformist, (informally) proves a metatheorem $\Phi(\mathcal{R})$, to the effect that the proofs in $\mathcal{R}$ suffice in principle for all the mathematics that C wishes to prosecute. (This is Metatheorem 2 above: any proof Θ that the conservative C produces has an austere counterpart $\rho\Theta$ in the reformist's system $\mathcal{R}$.) Let the most natural formalization, in $\mathcal{R}$, of R's informal proof of his metatheorem be called Π.
3. The reformist R agrees with the conservative C that many a result that C can prove in $\mathcal{C}$ enjoys, in $\mathcal{R}$, only infeasibly long proofs.

Both Burgess and I agree on 1, 2 and 3. But we disagree on further reflections that Burgess enters on the heels of 2. In the passage quoted above, the crucial complaint that Burgess makes is as follows:

> ... [$\mathcal{R}$] relies on a theorem ... for which only a classically and not a perfectionistically acceptable "proof" *has* been given. [My emphasis — NT]

In [2005b], Burgess expresses this objection in a slightly different fashion:

> The rules of the cut-free system cannot be the whole story about what guides usage. For if they were, then the cut rule — used in practice throughout intuitionistic as well as classical mathematics — would not be accepted until the cut elimination theorem had been given a cut-free proof. But all the proofs of that theorem that have actually been given use the informal counterpart of the cut rule. To be sure, these proofs are constructive, and implicitly provide a recipe for converting a proof with cut into one without, which recipe could in principle be applied to the very proof of cut elimination itself. But the recipe results in a superexponential increase in the length of proofs, *and would give a cut-free proof of cut elimination having a number of steps probably greater than the number of elementary particles in the visible universe.* In practice no proof this long ever will be given by anyone. [My emphasis — NT]

The debate between the reformist R and the conservative C now takes more definite shape. Burgess's critical view (on C's behalf) now consists of the claims 1, 2 and 3, but with the following rider to 2: **the austere formalization $\rho\Pi$ (in $\mathcal{R}$) of Π must be infeasibly long.**

This rider I deny, on behalf of the reformist R. Burgess offers no *proof* (by whatever the prevailing standards of rigorous but informal mathematical proof we are abiding by) for the claim that the specific proof in $\mathcal{R}$ of the metatheorem $\Phi(\mathcal{R})$ must itself be infeasibly long. In direct rebuttal, I offer the conjecture that its formalization, as a natural deduction *in normal form*, within a system of iterated inductive definitions, would actually be eminently surveyable. The reformist R is perfectly entitled to use such a proof system; it has been offered as the natural formalization of a significant fragment of intuitionistic mathematics. One must not confuse (as Burgess appears to have done)

(a) the known risk of exponential blow-up in length of object-system proofs upon normalization (or cut-elimination), the precise extent of which of course depends on the structural niceties of the particular abnormal proof that one is trying to normalize, with

(b) the unsubstantiated conjecture that the particular proof Π, at the metalevel, of the metatheorem $\Phi(\mathcal{R})$, would itself, upon normalization, become unfeasibly long.

This unsubstantiated conjecture is, I believe, unsubstantiable — indeed, refutable. A refutation would consist in simply producing a normalized proof $\rho\Pi$ of $\Phi(\mathcal{R})$, in the form of a user-friendly computer-printout of a proof-term, or a token of the natural deduction on a large enough webpage — up, down and across which the human verifier could scroll at leisure.

Metatheorem 2 follows easily from the fact that all proofs Σ of φ from Δ can be normalized (even if not in general *feasibly*), and then relevantized, so as to produce a proof $\rho\Sigma$, in the system $\mathcal{R}$, of either φ or $\perp$ from (some subset of) Δ. I believe that this transformability result admits of feasible proof, in normal form, in some appropriate (relevantized) metasystem $\mathcal{R}^*$ of iterated inductive definitions. In support of this contention I can only offer, at this stage, three admittedly inconclusive considerations.

First, *pace* Burgess, I do not see a plethora of cuts that need to be eliminated from within the usual inductive metaproofs of cut-elimination that are to be found in textbooks of proof-theory. They strike me as being already pretty close to normal form.

Secondly, such proof-theoretic intuitions as I have developed in frequent pursuit of fully formalized proofs *in normal form* of interesting mathematical theorems make me optimistic about the prospect of furnishing an $\mathcal{R}$-proof of the metatheorem.

Thirdly, there is the further consideration — not irrelevant or inconsequential in this setting — of the automation of proof-search. Fully automated proof-search proceeds in such a way that the output is always a proof in normal form. Computers could easily be recruited to the task of both producing and verifying a proof-term in normal form whose conclusion is $\Phi(\mathcal{R})$, should it turn out that the task were too much for the mental powers and digital dexterity of a human mathematician. In pursuit of this goal, one could legitimately resort to interactive theorem-proving, in order to increase one's chances of success. The proof arrived at would still have its (computationally verifiable) formal credentials unimpeached, despite the inputs from the human user that might have been practically necessary in order to enable its discovery.

Should C (or Burgess himself) scoff at this foreshadowed prospect of intellectual prosthetics, let me offer a *tu quoque*. Consider points 1–3 above with the following interpretation:

(i) $\mathcal{C}$ contains just the informal proofs, in mathematical English, of the kind that one finds in mathematical textbooks and journals. These are proofs in which one resorts to abbreviatory devices such as 'w.l.o.g.', 'by a similar argument', 'the remaining cases are dealt with similarly', and helps oneself liberally to 'single steps' of logical argumentations that are massively compiled, when considered as transitions within a formal logical system.

(ii) $\mathcal{R}$ contains only fully formalized natural deductions in the system *LK*, say, of Gentzen.

It is remarkable what blow-up in the length of 'proof'-tokens is involved in the simple (and supposedly — for Burgess — unproblematic) transition from $\mathcal{C}$-proofs to $\mathcal{R}$-proofs, for various quite simple mathematical theorems. So if computer-unaided feasibility were his *crie du jour*, Burgess could well be hoist with his own petard. For then the conservative C working 'in' the system $\mathcal{C}$ of *fully formalized* proofs (involving, to be sure, cuts galore) might well be a figment of the *classical* methodologist's imagination.

Whether one were to succeed by oneself, unaided by cybernetic deskmates, in writing down a normal-form proof of $\Phi(\mathcal{R})$, or needed to resort instead to automated proof-search and -verification in order to produce a token (soft or hard) of such a proof, the direct or indirect success involved would be an adequate dialectical counter to Burgess's assertion that **the austere formalization $\rho\Pi$ (in $\mathcal{R}$) of Π must be infeasibly long.**

Suppose, then, that a feasible austere proof $\rho\Pi$ is available for the metatheorem $\Phi(\mathcal{R})$. What would this enable the reformist R to maintain, philosophically or methodologically? I believe it would put R in a very interesting position, which deserves to be expounded briefly.

First, R could maintain that the system $\mathcal{R}$ really is the *correct* logic, on meaning-theoretic grounds, and on grounds of its (now feasibly!) proven adequacy for mathematics and science. Secondly, R could graciously concede point 3 above, and *draw his interlocutor's attention to the striking fact* that — despite $\mathcal{R}$'s really being the correct logic — honest and sincere deducers nevertheless resort to cut (i.e., to the interpolation of lemmas between their axioms and their theorems). *R can now offer a pragmatic explanation of this striking fact.* They reason using cut, according to R, because his very own metatheorem — feasibly proven! — shows that whatever they can prove using cut can be proved (austerely, and perhaps with epistemic gain) in $\mathcal{R}$, *even if only in principle.* If, in fact, it were to be infeasible, for particular $\mathcal{C}$-proofs, to render them as austere and possibly gainful $\mathcal{R}$-proofs, we now understand why reasoners resort to $\mathcal{C}$-proofs. For many a desired theorem, they *have* to, in order to get any semblance of the deductive job done. Moreover, since they are usually working within a system whose consistency is a well-tested article of faith, they know with that degree of certainty that any result proved by means of a $\mathcal{C}$-proof Θ will be true in all interpretations of the axioms used in Θ.

In an analogous way the pragmatist nominalist can explain why we incur commitment to numbers as theoretical objects. It is because their introduction, into our ontology, by the linguistic devices of numerical reference and quantification enables us to prove, very efficiently, empirical predictions from empirical premises whose number-eschewing proofs would be tortuously long. Undertaking these Platonistic commitments enables the mathematizing reasoner to furnish shorter and often more elegant proofs of predictions from his physical theory. This is the central idea of Hartry Field's [1980]. It is worked out in considerable detail via conservative extension results. These aim to show that grafting mathematics onto a 'synthetic' theory enables one to prove only such synthetic claims from synthetic premises as are provable (albeit much more laboriously) *within* the synthetic theory *without* resorting to the deductive shortcuts afforded by the mathematics.

Similarly, I have argued, in [Tennant, 1996],[5] that when R is the intuitionist and C (as usual) is the classicist, R can give a pragmatic explanation of why C resorts to strictly classical inferential moves not countenanced within intuitionistic logic. Such moves are explained as expressive of high-level, *realist*, metaphysical commitments — commitments to the *determinacy of the world* in the contextually relevant regards. Undertaking these commitments enables the classical reasoner to furnish shorter and often more elegant proofs of results that the intuitionist would be able to prove only in a much more laborious (albeit more informative) way. Moreover, the intuitionist can prove, intuitionistically, an impressive array of results of the general form 'If there is a $\mathcal{C}$-proof of $\perp$ from Δ, then there is an $\mathcal{R}$-proof (i.e., an intuitionistic proof) of $\perp$ from Δ'. These are known as *relative consistency theorems*, and the important point is that they are *intuitionistically provable*. They are the intuitionist's way, from within the 'limitations' of his own vantage point, of explaining why no great epistemic disaster is going to befall the classical mathematician who uses the full system $\mathcal{C}$ of classical logic in order to reach results that have no constructively acceptable warrant for assertion.

---

[5] Burgess appears, in his review [2005b] of *The Taming of The True*, to be unaware of this aspect of my overall view.

We see, then, that the 'austere eschewer' R — whether he be a nominalist eschewing numbers, or an intuitionist eschewing classical moves of inference, or a relevantist ('perfectionist') eschewing cuts, *ex falso quodlibet* and kindred irrelevancies or potential sources of irrelevance — has a uniform way of explaining why it is that the conservative opposition incurs commitment to numbers, reasons classically about them, and uses cut when doing so. The 'philosophically suspect' maneuvers are all made in the name of greater efficiency and greater scope in our mathematical and scientific theorizing. The challenge is to find an austere vantage point from which one can explain, in austere terms, how this has come to pass; and from within the 'limitations' of which one can show that no great epistemic disaster will befall those who are tempted to go beyond the austere canons $\mathcal{R}$ of correct reasoning, and to adopt the more expansive system $\mathcal{C}$ as a methodologically optimizing extension. This extension is motivated by a firm belief in the consistency of mathematics; a commitment to abstracta; and a realist conviction that the world is determinate in various regards. The challenge, to repeat, is to make austere sense of all this. And I do not believe that Burgess's unsubstantiated conjecture about the feasibility of a particular metaproof is an insuperable obstacle to doing so.

BIBLIOGRAPHY

[Bridges and Richman, 1999] D. Bridges and F. Richman. A constructive proof of Gleason's Theorem. *Journal of Functional Analysis*, 162:287–312, 1999.

[Bridges, 1999] D. Bridges. Can constructive mathematics be applied in physics? *Journal of Philosophical Logic*, 28(5):439–453, 1999.

[Burgess, 2005a] J. Burgess. No Requirement of Relevance. In Stewart Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 727–750. Oxford University Press, New York, 2005.

[Burgess, 2005b] J. Burgess. Review of Neil Tennant, *The Taming of The True. Philosophia Mathematica*, 13:202–215, 2005.

[Dummett, 1991] M. Dummett. *Frege: Philosophy of Mathematics.* Harvard University Press, Cambridge, Massachusetts, 1991.

[Field, 1980] H. Field. *Science Without Numbers: A Defence of Nominalism.* Basil Blackwell, Oxford, 1980.

[Hellman, 1993] G. Hellman. Gleason's Theorem is not constructively provable. *Journal of Philosophical Logic*, 22:193–203, 1993.

[Quine, 1990] W. V. O. Quine. *The Pursuit of Truth.* Harvard University Press, Cambridge, MA, 1990.

[Tennant, 1985] N. Tennant. Minimal logic is adequate for popperian science. *British Journal for Philosophy of Science*, 36:325–329, 1985.

[Tennant, 1987] N. Tennant. *Anti-Realism and Logic: Truth as Eternal.* Clarendon Library of Logic and Philosophy, Oxford University Press, 1987.

[Tennant, 1992] N. Tennant. *Autologic.* Edinburgh University Press, 1992.

[Tennant, 1994] N. Tennant. Intuitionistic mathematics does not need *ex falso quodlibet*. *Topoi*, pages 127–133, 1994.

[Tennant, 1996] N. Tennant. The Law of Excluded Middle is Synthetic A priori, if Valid. *Philosophical Topics*, 24:205–229, 1996.

[Tennant, 1997] N. Tennant. *The Taming of The True.* Oxford University Press, 1997.

DEFAULT REASONING

Nicholas Rescher

1 DEFAULT INFERENCE

The topic of default reasoning affords some instructive insights into the nature of the relationship between theoretical and practical reasoning. A *default* in logic is a fall-back position in point of conclusion-drawing — one to which we can appropriately take resort when things go wrong. But of course things ought not to go wrong in logic. So what is going on here?

Orthodox inferential reasoning proceeds via *logically valid* inference processes which, as such, do — and must — lead to true conclusions when the premises are true. By contrast, default reasoning — which involves an information-gap between premises and conclusion — is such that premises that are true will lead to plausible (though possibly false) conclusions.

The logical validity of inference rules in standard (truth-functional) logic is determined on an input-output basis, a valid rule being one that will invariably yield true outputs (conclusions) from true premisses. All such inference rules will faithfully and unfailingly transmit the truth of premisses to the conclusions. By contrast, the inference processes of default logic are such that the truth of the premisses does not assure that of the conclusion but will at most establish that conclusion as plausible. Such inferences are *ampliative*: the conclusion can go beyond what the premisses provide, thanks to a shortfall of information. And this means that such reasonings are fallible and can — and occasionally will — lead from true premises to false conclusions.

We shall represent *logically valid deducibility* (in its classical construction) by $\vDash$, and by contrast use $\vDash$ to represent the *plausible inferability* at issue with default reasoning.

Some examples of inference-processes in default logic are as follows:

1. p is highly likely $\vDash p$

2. p is very likely, q is very likely $\vDash p\&q$ is very likely

3. there is strong evidence in favor of p and no more than weak evidence against it $\vDash p$

4. p has obtained in all past instances $\vDash p$ will obtain in the next instance

Handbook of the Philosophy of Science. Philosophy of Logic
Volume editor: Dale Jacquette
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

If all we are told of some number is that it is a prime, we would, plausibly enough, conclude that it is not an even integer — even though we are aware that this conclusion will prove false once out of an infinity of cases (viz. that of the number two).

As these examples indicate, the inference processes of default logic can all be assimilated to a deductive pattern of the following structure (which does clearly obtain as valid in traditional logic):

- In all ordinary (normal, standard, commonplace) cases, whenever P, then Q.
- P obtains in the case presently at hand
- ⟨The present case is an ordinary (normal, standard, etc.) one.⟩

$\therefore$ Q obtains in the present case

Here that third, usually tacit and thereby enthymematic, premiss plays a pivotal role. And it is, in general, able to do so not because we have secured it as a certified truth, but simply because it is a plausible (albeit defeasible) presumption that is strongly supported by the available evidence; though it is not, of course, guaranteed. Default reasoning accordingly rests on arguments which would be valid if all of their premisses — explicit and tacit alike — were authentic truths, which they are not, since at least one of the critical premisses of the argument is no more than a mere presumption.

Such a *defeasible presumption* is emphatically not to be regarded as an established truth, but merely something that holds only provisionally, as long as is no counter-indicatively conflicting information comes to light. Against this background the procedure that is definitively characteristic of default reasoning is:

> To treat what is generally (normally, standardly, generally, usually, etc.) the case *as if* it were the case always and everywhere, and therefore as applicable in the present instance.

Here, in effect, ignorance is bliss: where there is no good reason to see the case at hand as being out of the ordinary, we simply presume it to be an ordinary one in the absence of visible counter-indications. Such reliance on a *principle of presumption* to the effect that what generally holds also holds here in the case presently at hand that defines the *modus operandi* of default reasoning.[1]

2 FACING THE PROSPECT OF ERROR

Of course such plausible presumption can go awry. For it may well happen that the situation at hand fails to be standard and representative as the enthymematic comportment of the argument requires. This is brought out vividly in John Godfry

---

[1] Regarding default reasoning and its ramifications see [Horty, 2000; Harper, 1981; Kyburg and Teng, 2001; Rott, 2001; Bochman, 2001; Pollock, 1991; Reiter, 1987; Rescher, 1980].

Saxe's poem "The Blind Men and the Elephant" which tells the story of certain blind sages, those "six men of Indostan/To learning much inclined/Who went to see the elephant/(Though all of them were blind)." One sage touched the elephant's "broad and sturdy side" and declared the beast to be "very like a wall." The second, who had felt its tusk, announced the elephant to resemble a spear. The third, who took the elephant's squirming trunk in his hands, compared it to a snake; while the fourth, who put his arm around the elephant's knee, was sure that the animal resembled a tree. A flapping ear convinced another that the elephant had the form of a fan; while the sixth blind man thought that is had the form of a rope, since he had taken hold of the tail.

> And so these men of Indostan,
> Disputed loud and long;
> Each in his own opinion
> Exceeding stiff and strong:
> Though each was partly in the right,
> And all were in the wrong.

None of those blind sages was altogether in error; it is just that the facts at their disposal were nontypical and unrepresentative in a way that gave them a biased and misleading picture of reality. It is not that they did not know truth, but rather that an altogether plausible inference from the truth they knew propelled them into error.

But since such a policy of typicality presumption may well lead us down the primrose path into error, how is it ever to be justified? The answer here lies precisely in the consideration that what is at issue is not a truth-claim but a policy or procedure. And such policies of procedure are not justified in the theoretical (i.e., factual) order but in the practical or pragmatic order of deliberation. The validation at issue runs roughly as follows:

1. We have questions to which we need a (satisfactory) answer, and in the face of this we take the stance that —

2. We are rationally entitled to use a premiss that holds good promise of finding one (i.e., is effective or more effective than the other available alternatives) even though it may occasionally fail.

On this basis we proceed subject to the idea that if and when things go wrong, this is a bridge we can cross when we get there, invoking "explanations" and excuses to indicate the unusual (anormal, extraordinary) circumstances of the case.

Even as in real life we cannot manage our affairs sensibly without running risks, so in cognitive life one must, on occasion, take the risk of error in stride, since the inevitable result of a radical nothing-risk policy is the nothing-have of radical skepticism. And this situation is particularly prominent in inductive contexts.

3 INDUCTION AS DEFAULT REASONING

The term "induction" is derived from the Latin rendering of Aristotle's *epagôgê* — the process for moving to a generalization from its specific instances.[2] Gradually extended over an increasingly wide range, induction can be seen as a question-answering device encompassing virtually the whole range of non-deductive reasoning. Thus consider a typical inductive argument — that from "All the magnets we have examined attract iron filings" to "All magnets attract iron filings." It would be deeply problematic to regard this as a deductive argument that rests on the (obviously false) premiss: "What is the case in all examined instances is universally the case". Rather, what we have here is a plausible presumption that takes the cases in hand to be typical and generally representative in the absence of concrete counterindications — that is, an instance of default reasoning.

Induction, so regarded, is accordingly not so much a process of *inference* as one of presumption-based *truth-estimation.* We clearly want to accomplish our explanatory gap-filling in the least risky, the minimally problematic way, as determined by plausibilistic considerations. This is illustrated by such examples as:

- There is smoke yonder
- Usually, where(ever) there's smoke, there's fire
- ⟨The present situation fits the usual run⟩

∴ There is fire yonder

or again

- Two thirds of the items in the sample are defective
- ⟨The sample is representative of the whole⟩

∴ Two thirds of the items in the whole population are defective

(Here the enthymematically tacit premisses needed to make the argument deductively cogent have been indicated.)

Its reliance on a presumption of typicality, normalcy, or the like, means that any inductive process is inherently chancy. Induction rests on presumption-geared default reasoning and its conclusions are thus always at risk to further or better data since what looks to be typical or representative may in due course turn out not to be so.

4 DEFAULT REASONING AS NONMONOTONIC

In virtue of the fact that default reasonings rest on a presumption of normality, typicality, or the like, it may well transpire that while a premises implies a certain conclusion nevertheless the conjunction of this premise with some further propositions may fail to do so. Such implications are called non-monotonic because while

---

[2] See [Ross, 1949, 47-51].

"If p then q" obtains, nevertheless it can happen that q sometimes fails to obtain in certain circumstances, where p holds, so that:

$$p \Rightarrow q \text{ need not yield } (p \& r) \Rightarrow q$$

Additional information can destabilize default implications.

Clearly, the reason why the monotonicity-characterizing principle

- Whenever $p \Rightarrow q$, then $(p \& r) \Rightarrow q$

works in deductive context, is that here there is then no normality linkage between p and q which requires the addition of further material that may or may not be forthcoming as per a stipulation of normalcy or of "all things equal" in the case of inductive reasonings. The reliance of default reasoning on a presumption of normality, typicality, or the like, means that throughout this domain new information can undo earlier findings.

Thus consider the claim

— If you are in America, then you might be in New York.

This is, of course, perfectly correct. But it will not do to "strengthen" the antecedent as per

— If you are in America and you are in Texas, then you might be in New York.

The conclusions we arrive at with nonmonotonic implication relations are no more than presumption. For in making the inference we have to presume that the situation is not one where some yet unseen conclusion-averting circumstance comes into operation.

This state of affairs also means that with nonmonotonic implications *modus ponenes* fails: the combination of p and $p \Rightarrow q$ need not *demonstrate* that q obtains but may do no more than to establish a *presumption* to that effect.

Nonmonotonicity is thus a standard feature of default inference as is illustrated by contrasting

— If I had put sugar in the tea then it would have tasted fine

with

— If I had put sugar and cayenne pepper in the tea, then it would have tasted fine.

Or again, contrast:

— If you greet him, he will answer politely.

with

— If you greet him with an insult, he will answer politely.

After all, that first implication effectively (but tacitly) comes to

— If you greet him *in the usual and ordinary way*, he will answer politely

and the antecedent of the second implication violates that initial condition.

With default inferences we have to do with what is, from the standpoint of standard logic, a decidedly eccentric mode of reasoning. For no qualification additional to the antecedent as such can abrogate what a valid monotonic implication implies: the antecedent will, in and of itself, suffice to guarantee the consequent. But whenever that "inevitably (invariably unavoidably, etc.)" becomes weakened to "generally, usually, probably, possibly, etc.)," the monotonicity that is requisite for authentic implication is lost. To obtain a conclusion we must now suppose that nothing untoward is hidden from our sight — that nothing unmentioned intervenes. And this always brings the factor of presumption upon the scene.

5 SOME COMFORTING CONSIDERATIONS

But what if those normality presumptions should prove unjustified? How are we to proceed in the context of conclusions arrived at by reasoning that we see as potentially misleading? The short answer is: Cautiously! But a somewhat more informative response lies in the important prospect of *blurring* that conclusion — making it less specified and detailed. As stated at the outset default reasoning calls for the possibility of resort to a fall-back position. And in managing our cognitive risk's we can always fall back upon *vagueness* and its inherent qualifications.

With default reasoning in general and induction in particular we run the risk that our conclusions may run awry thanks to our reliance on (generally tacit) suppositions of normality or typicality that may fail in the circumstances at hand. To offset the risk error we can resort to the introduction of decreasing definiteness for the sake of increasing security. Thus instead of reasoning

- q is highly likely wherever p
- In the present case p obtains
- The present case looks to be a typical, majority-conforming one
  ⟨Looks are not deceiving here⟩

∴ In the present case q obtains

we would instead reason to

- In the present case q probably obtains

Thereby taking a sensible step in the direction of safety. But of course likelihoods do not answer yes/no questions, and where such questions confront us we have little choice but to resort (circumstances permitting) to chance the risks of the presumption of typicality/normality that characterizes default reasoning. There are, however, some promising precautions here.

After all, a fundamental feature of inquiry is represented by the following observation:

THESIS 1. Insofar as our thinking is vague, truth is accessible even in the face of error.

Consider the situation where you correctly accept P-or-Q. But — so let it be supposed — the truth of this disjunction roots entirely in that of P while Q is quite false. However, you accept P-or-Q only because you are convinced of the truth of Q; it so happens that P is something you actually disbelieve. Yet despite your error, your belief is entirely true.[3] Consider a concrete instance. You believe that Mr. Kim Ho is Korean because you believe him to be a North Korean. However he is, in fact, a South Korean, something you would flatly reject. Nevertheless your belief that he is Korean is unquestionably correct. Thanks to the indefiniteness of that disjunctive belief at issue, the error in which you are involved, although real, is not so grave as to destabilize the truth of your belief.

This example illustrates a more far-reaching point.

THESIS 2. There is, in general, an inverse relationship between the precision or definiteness of a judgment and its security: detail and probability stand in a competing relationship.

It is a basic principle of epistemology that increased confidence in the correctness of our estimates can always be purchased at the price of decreased accuracy. We *estimate* the height of the tree at around 25 feet. We are *quite sure* that the tree is 25±5 feet. We are *virtually certain* that its height is 25±10 feet. But we are *completely and absolutely sure* that its height is between 1 inch and 100 yards. Of this we are completely sure, in the sense that we deem it absolutely certain, secure beyond the shadow of a doubt, as certain as we can be of anything in the world, so sure that we would be willing to stake our life on it, and the like. With any sort of estimate, there is always a characteristic trade-off relationship between the evidential *security* of the estimate on the one hand (as determinable on the basis of its probability or degree of acceptability), and the informative *definiteness* (exactness, detail, precision, etc.) of its asserted content on the other. Vaguer and looser statements are for that very reason more secure because they embody larger margins of error. This relationship between security and definiteness is graphically characterized by a curve of the general form of an equilateral hyperbola. (See Display 1.) and this sort of relationship holds just a well for our *truth* estimates as of others.

This state of affairs has far-reaching consequences. It means, in particular, that no secure statement about objective reality can say exactly and in complete detail how matters stand universally always and everywhere. To capture the full complexity of the truth of the matter of things by means of language we must often proceed by way of "warranted approximation." In general we can be sure of how things "usually" are and how they "roughly" are, but not how they always and

---

[3] Examples of this sort indicate why philosophers are unwilling to identify *knowledge* with *true belief*.

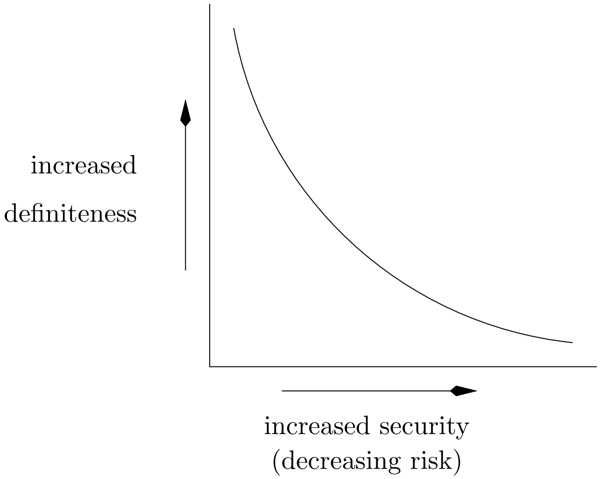


Figure 1. The Security/Definiteness Relationship
Note: The overall quality of the information provided by a claim hinges on combining its security with its definiteness. Given suitable ways of measuring security (s) and definiteness (d) the curve at issue can be supposed to be an equilateral hyperbola obtained with $s \times d$ as constant.

exactly are. And this impels our reasoning in the direction of presuppositions of normalcy, typicality, and the like, which are characteristic of default argumentation.

But be this as it may, the present considerations indicate that “inductive inference” as traditionally conceived affords a paradigm instance of default reasoning, which itself emerges in their light as an exercise in standard deductive inference subject to a recourse to the potentially defeasible presumption of typicality.

Yet how is the adoption of a potentially defeasible thesis to qualify as rationally appropriate? The answer, as noted above, lies in the general principle of risk management. For what is at issue with presumption is at bottom less an endorsement of the truth than the implementation of a policy. And rationality here — as elsewhere in matters of practical procedure — pivots on the principle of a favorable balance of potential benefit over potential loss. In many situations default reasoning affords our best-available pathway to our ultimately very practical need for information — for answering in a cogent and epistemically responsible way a question that we need to resolve. For in truth-estimation as in so much of life we have to rest content with doing the best one can actually manage to achieve in the circumstances.

BIBLIOGRAPHY

[Bochman, 2001] A. Bochman. *A Logical Theory of Nonmonotonic Inference and Belief Change.* Berlin: Springer Verlag. 2001.

[Harper, 1981] W. L. Harper. A sketch of some recent developments in the theory of conditionals. In Harper, L.G. Pearson and Robert Stalnaker, eds. *IFS: Conditionals, Belief, Decision, Chance and Time*. Dordrecht: D. Reidel. 1981.

[Horty, 2000] J. Horty. Common-sense reasoning. *The Routledge Encyclopedia of Philosophy.* London: Routledge. 2000.

[Kyburg and Teng, 2001] H. E. Kyburg and C. M. Teng. *Uncertain Inference*. Cambridge: Cambridge University Press, 2001.

[Pollock, 1991] J.L. Pollock. A theory of defeasible reasoning. *International Journal of Intelligent Systems*, 6, 33-54. 1991.

[Reiter, 1987] R.B. Reiter. Nonmonotonic reasoning, *Annual Review of Cognitive Science*, 2, 147-186. 1987.

[Rescher, 1980] N. Rescher. *Induction*. Oxford: Basil Blackwell. 1980.

[Ross, 1949] W.D. Ross. *Aristotle's Prior and Posterior Analytics*. Oxford: The Clarendon Press. 1949.

[Rott, 2001] H. Rott. *Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford: The Clarendon Press. 2001.

INDEX

T^F, 1100
$\mathfrak{A}$-normal structures, 901
0-Model, 1070

Abar, C. A. A. P., 872
abductive inference, 1116
Abe, J. M., 874, 903
Abelardian, 1098
absolute, 633–634
absolute infinity, 525, 526, 533–536, 545
abstract entity, 454, 479
abstract fuzzy deduction system, 712
abstract fuzzy semantics, 713
abstract objects, 423
abstraction, 484
accessibility relation, 927–929, 944, 946, 954
accuracy, 58
Ackermann, W., 57, 208, 415, 417, 418, 419, 425, 431, 436, 441, 500–503, 511, 519, 723, 725, 726, 772, 773, 778
  Ackermann function, 513, 772, 778, *see* complexity
  Ackermann property, 773
actual infinity, 523–548
actual world, 1099
actuality operator, 967
actuarial reasoning, 164
Aczel, P., 390, 517, 529
Adams, F., 1077
adjectivism, 1105
adjointness condition, 698
admissible rule, 738, 773, 946
  γ, 773
  cut, 738, 739, 741, 756, 760, 761, 781, 784
adverbilization, 1105
Aegean Sea, 160
Agatha Christie, 1075
Ajdukiewicz, K., 921
Akama, S., 874
Akiba, K., 1091
Alcântara, L. P., 872
$\aleph_1$-like ordering, 599, 606
$\aleph_\alpha$, 530
alethic modality, 1098
algebra of confidence, 890
algebraic methods of formal logic, 134
algebraic methods of mathematical logic, 135
algebraic semantics, 218, 686
algorithmic compleixty, 169
algorithmic methods of mathematical logic, 135
$\forall$, constructive interpretation, 554–557
Allis, V., 572
allowable inconsistencies, 1079
Allwein, G., 777
α-model, 680
alteration, 1134
alternative set theory (AST), 528, 562
Alves, E. H., 828, 870
Alward, P., 1061, 1085, 1117
Alzheimer illness, 874
ambiguity, 159, 1064
analysis, 142
analytic tableaux, 729, 777, 781
  for **fde**, 729
  for classical logic, 729
$\wedge$, constructive interpretation, 553–555

Anderson, A. R., 208, 220, 346, 375, 723, 724, 726, 729, 731, 732, 746, 747, 773, 775, 780
Andrews, P. B., 923
Angel, L., 577
annotated atoms, 859, 886
annotated clause, 867
annotated logic, 796, 858, 859, 863, 867, 874, 876, 886
  inductive, 884
Anscombe, G. E. M., 44, 143
Anselm, 147, 149–152
  Anselm's ontological argument, 146, 149
applied logic, 41, 43, 793
appreciator make-believe, 1083
Åqvist, L., 1069
Arai, T., 435, 436
Archimedes, 46
argument, 132
  analysis, 122–123
  complex, 108–109
  construction, 126
  convergent, 122
  criticism, 126
  disputational sense, 102
  evaluation, 124–125
  evidence-accumulating, 123
  forms, 47, 58
  identification, 121
  linked, 122
  reason-giving sense, 102
  simple, 107–108
  subordinate, 109
  suppositional, 110–114
Argument T, 1003
argumentation, 118
argumentation scheme, 124, 125
Aristotle, 41, 42, 45, 47, 133–135, 166, 170, 523, 535, 567, 570, 681, 943, 958
  Aristotelian logic, 43–45, 57, 131, 135, 792
arithmetic, 50
arithmetisation of analysis, 527
Armour-Garb, B., 1079
Armstrong, D., 980
Arruda, A. I., 792, 804, 808, 810, 837, 870, 872
artificial dependence, 1137
artificial intelligence, 208, 858, 875, 884
As-Sīrāfī, 44
assignment, 590
asymmetry, 1093, 1095
ATMS model, 1078
Auerbach, D., 435
Aurifaber, J., 48
Austin, J. L., 1069
author's sayso axiom, 1068, 1069, 1078, 1082, 1092, 1093, 1094, 1096
Avigad, J., 414, 418, 437, 440
Avila, B. C., 874
axiom of choice, 497, 516–518, 591
axiom of foundation, 517
axiom of infinity, 512
axiom of reducibility, 463, 481, 510
axiom of separation, 516
axiomatic method, 411, 412
axiomatic calculus, 723
  relevance logics formalized as, 733, 736, 740, 754, 756, 761, 762, 774, 777, 779, 780, 784
axioms, 50–52
Azzouni, J., 1107

B (combinator), 733, 734, 738, 748, 751
  simple type of, 733, 745, 754, 776
B' (combinator), 733, 734, 736–738, 748, 750, 751, 757, 780
  simple type of, 733, 745, 747, 754, 779
B (basic relevance logic), 774–775
Béziau, J. -Y., 871, 903
background assumption, 1112
backtracking, 1130, 1131, 1136
backward causation, 1130

bad arguments, 42, 43, 47, 59
Baire, R., 531, 552
Balaguer, M., 531
balanced coherence frame, 981
balanced standard frame, 982
Bambrough, R., 167, 168
Banach, S., 568
Banach–Tarski paradox, 497
Bar-Hillel, Y., 921
barbara, 45, 49
Barcan, R. C., 958
Barcan formula, 930, 960, 1073
Barconists, 1105
Barreto, M. M. G., 875, 903
Barwise, J. K., 728
Barwise, J., 41, 137, 389–391, 607, 608
basic t-norm logic, 703
Batens, D., 837, 873, 1079
Bayesian probability, 890
Bays, T., 364
Bazhanov, A. A., 793
Beall, JC, 343, 361, 368, 373, 375–377, 402, 403
Beardsley, M. C., 122, 133
Becker's postulate, 147
Beeson, M. J., 548, 554–557, 564
Begriffsschrift, 50–52, 498
Behmann, H., 414, 415
belief revision, 68
belief systems, 857
Bell, J. L., 527, 528
Belnap, N. D. Jr., 208, 224, 339, 375, 392, 395, 396, 723, 724, 726, 729, 731, 747, 755, 773–775, 777, 780, 781, 784
Benacerraf, P., 468, 527, 571, 572, 619, 620
  Benacerraf's Dilemma, 1061, 1062
Benardete, J. A., 525, 565, 567, 570, 574, 576
Bencivenga, E., 147, 1077
van Benthem, J. F. A. K., 966
Bernays, P., 411, 415, 417–419, 421–426, 431, 432, 441, 449, 463, 501, 503, 505, 518, 543, 551, 564
Bernays-Schönfinkel classes, 610
Bernoulli, J., 840
bet-sensitive, 1070, 1096
Beth operation, 599
Bibel, W., 876, 879
bilattice, 697, 858
Bimbó, K., 739, 751, 778
binary opartions, 218
Birkhoff, G., 794, 892
Bishop, E., 546, 548, 554, 558, 563, 564
Bishop, M. A., 164
BIV, *see* Bivalence
bivalence, 327, 328
bivalent truth value semantics, 137
BL-algebra, 703
Black, M., 162, 163, 572
Blackburn, S., 403, 926
Blair, H., 858, 874
Blair, J. A., 101, 119
Blau, U., 689, 692
Blocker, H. G., 1061
Bloo, D. da, 1103, 1115
blueprint strategy, 1139, 1141, 1142
Bočvar, D. A., 682, 689
Bocheński, J. h M., 943
Boethius of Dacia, 41, 43
Bohr, N., 792, 893, 894, 896
  the Como lecture, 894
Bolzano, B., 42, 43, 53, 527–529, 964
  Bolzano's Program, 76
bookkeeping axioms, 711
Boole, G., 41, 43, 48, 49, 497
Boolean algebras, 133, 890, 949
Boolos, G., 441, 542–543, 746, 929
Born, M., 795
bottom-up theories of constructive proof, 555–556, 558
Boukreev, A., 164
boundaries, physical, 568

bounded, 605
bounded occurrence, 602
Bourbaki, N., 54, 840, 849
box distribution, 967
Bradbury, R., 1069, 1095, 1100, 1115
Brady, R. T., 724, 747
brain, 42, 51, 58
Brandom, R., 367
Brecher, R., 147
Brett, J., 1075, 1083
Bridges, D., 548, 554, 564
Brouwer, L. E. J., 165, 411, 415, 416, 498, 525, 531, 546, 547, 550, 553, 564, 794, 944, 953
Brown, B., 1079
Brown, M., 957
Buchsbaum, A. T. C., 875
Bueno, O., 857
Bunder, M., 871, 873
Bunge, M., 1104
Burali-Forti's paradox, 843
Burge, T., 341, 386, 387, 404
Burley, W., 47
Byrne, A., 1087, 1102

C (combinator), 736, 750, 760
  simple type of, 776, 779, 780
Caiero, R. da C., 831
Caldon, P., 435, 1110
can do principle, 1080
can't-help-it realism, 1070, 1071, 1073, 1105
canonical algebra, 949
Cantini, A., 405
Cantor, G., 412, 726, 413, 497, 498, 516, 518, 525, 527–546, 551, 727, 986
  Cantor's diagonal argument, 539, 540
  Cantor's finitism, 534
  Cantor's generating principles for sets, 539
  Cantor's naïve set theory, 830
  Cantor's paradox, 843
  Cantor's theorem, 616
Cao, T. Y., 578
cardinal numbers, transfinite, 529–530
cardinality quantifier, 598
Carlson, G. N., 1112
Carnap, R., 160, 161, 207, 457, 476, 504, 918, 919, 924, 925, 928, 931, 937, 939, 940, 958, 964
Carnielli, W. A., 870, 872, 903
Carroll, L., 133, 168
cartographers, 161
Cartwright, N., 1076
Casadio, C., 921
Castãnada, H.-N., 1075
categorial grammar, 921
categorical
  second order analysis, 900
  second order arithmetics, 900
  propositions, 829
Cauchy, A.-L., 840
causal
  chain, 1132
  theory, 1068
  theory of reference, 960
causality, 479
causation, 1127, 1131–1133
  as influence, 1134, 1135, 1140
central assumption, 1002–1004, 1009, 1011–1012, 1017, 1021
Cervantes, M., 1075, 1088, 1094, 1113
chaining, 121
Chang, C. C., 601
Chapuis, A., 405
characterise system E, 215
characteristic for a system, 220
Characterization principle, 151
Chellas, B. F., 354, 926
Chesterton, G. K., 157, 158
Chihara, C., 947
choice sequences, 563
Chomsky, N., 60, 560
Chuaqui, R., 851

- Church, A., 53, 54, 411, 498, 511–513, 519, 723, 727, 735, 749, 779, 781, 812, 835, 914, 920, 922, 935, 936, 958, 959
  - Church's thesis, 513, 514, 516
- Clarke, B. L., 567
- class, 48, 481
  - as a basis for sets, 536–537
- classical bivalent truth value, 140
- classical logic, 156, 165, 171, 206, 208, 212, 213, 723, 725, 729, 1067, 1068, 1080
  - in *curricula*, 723
  - in relevance logics, 723, 769, 773, 784
- classical logicians, 1103
- classical predicate logic, 213, 214
- closure, 1075
- closure-condition, 1075
- cognitive economy, 1111
- cognitive science, 42, 52, 59
- Cohen, P. J., 518, 540
- coherence, 1063, 1065, 1066
- Coleman, F. X. J., 1061
- collective belief world, 1087, 1100
- Collins, J., 1127, 1135
- colloquial, 142
- combinator, *see* B, B', C, I, K, M, S, S', W
- combinatorial view of sets, 543–544
- combinatory logic, 723, 727, 749, 779
- common experience, 1112
- common knowledge, 89, 1111
- communication, 68
- community-acceptance, 1105
- compact, 602
- compact sets, 524
- compactness theorem, 499, 507, 519
- comparative similarity, 1128
- competence-performance distinction, 560–561
- complementary
  - literals, 880
  - propositions, 896
  - theories, 894, 896
- complete, 1092
- complete world, 1112
- completeness, 411, 415, 500, 502, 503, 506, 507
  - of modal logics, 743, 757
  - of propositional logic, 415
  - of relevance logics, 731, 745, 746, 748, 760–762, 767, 768, 775, 778, 781, 782
- complex literals, 886
- complexity, 612, 730, 772, 778
  - of validity in **fde**, 730
  - of validity in classical logic, 730
  - primitive recursive in the Ackermann function, 772
- compositionality, 913, 915, 916, 925, 934
- comprehension axiom, 517, 1092
- comprehension principle, 531–532
- computability theory, 42, 515
- computational limits, 59
- concept, 915–918, 923–925, 928, 936, 937, 940
- conceptual analysis, 1063–1066, 1076, 1081, 1082, 1099, 1110
- condensed detachment, 748, 753, 754
- conditional, 147
  - logic, 78
  - proof, 163
- confidence function, 890
- Coniglio, M. E., 872
- connectionist, 170
- connections, 881
- connective
  - external, 690
  - internal, 690
  - strong, 690
  - weak, 691
- consecution calculus
  - $\mathbf{LR}_{+}^{\circ t}$, 737, 784
  - $\mathbf{LT}_{+}^{\circ t}$, 754–758, 760, 761
  - $\mathbf{LT}_{\rightarrow}$, 741
  - $\mathbf{LT}_{\rightarrow}^{\circ t}$, 739–740, 742, 755

  relevance logics formalized as, 737, 739–742, 755, 760, 761, 768
consequence, 1067, 1075, 1112
  abductive, 1112
  deductive, 1112
  inductive, 1112
  plausibilisitc, 1112
  presumptive, 1112
  statistical, 1112
  relations, 1112
conservative extension, 740, 741, 746, 756, 762, 781, 784
conservativity, 430–431
consistency, 1092
  and inconsistent multiplicities, 533
  minded theory, 435
proof, 412–415, 417–419, 425, 428, 430–437, 439, 441, 508
  theory, 791, 797
constants, 589
constitutive properties, 1092
constructibility structures, 1021
constructibility theory, 1021
constructible hierarchy, 518
constructive logic, 953
constructive mathematics, 794
constructivism, 525, 546
context principle, 459
contextualization, 1093
contingentia futura, 681
continuity, its role in supertasks, 571–572
continuum hypothesis, 498, 516
continuum, the structure of, 566–567
contradiction, 141–145, 147, 149, 150, 417, 797, 809, 1069, 1079
  -affirming, 1079
  data base, 215
  intended, 1079 subitem ontic view of, 899
  real world, 873, 899
  -tolerant, 1079
  truth, 899
contraposition, 163, 1129, 1132
conventionalism, 857
converse Barcan formula, 960
Cooke, M. C., 575
Copeland, J., 927
cops and robbers, 1070, 1085, 1086
coriander, 42
correct logic, 205, 206, 217
correct system, 208
correspondence theory, 454
Costa Leite, A., 903
Costa, N. C. A. da, 1079
Cotogno, P., 571
countably compact, 602
counterfactual, 78, 956, 1127–1130, 1132, 1133
  analysis of causation, 1132
  dependence, 1131, 1133, 1134, 1141, 1142
counterpart
  frame, 968
  structure, 970
  theory, 929, 1000
counting algorithm, as a basis for arithmetic, 548–552
creative authority, 1068
Cresswell, M. J., 926, 939
critical reasoning, 134, 135
Crittenden, C., 1061
cube lattice, 221
cumulative hierarchy of sets, 542
Currie, G., 1074, 1080, 1082, 1087, 1102
Curry, H. B., 723, 727, 733, 738, 749, 779
  Curry's paradox, 345–347, 351, 356, 357, 368, 373, 374, 378, 379, 398, 830
cut (rule), 738, 741, 756, 784
  admissible, 739, 743, 760, 761, 781
  single and multiple, 738
cylindrification, 965

D (rule), *see* condensed detachment
D'Ottaviano, I. M. L., 871
da Costa, N. C. A., 793, 844, 851, 857, 870–872
Dalla Chiara, M. L., 840, 849–851
data bases, 208, 211, 212, 214, 217
Dauben, J. W., 528
David, M., 403
Davidson, D., 478
Davies, E. B., 571
Davis, M., 513
Dawar, A., 611
D-completeness, 751–754
  of $\mathbf{T}_{\rightarrow}$, 754
  of BCI, 776
de dicto/de re distinction, 958, 1089, 1107, 1108
  de re-ification, 1073
de facto dependence, 1135–1138
De Morgan, A., 497
  De Morgan laws, 871
  De Morgan monoid, 782
  De Morgan negation, 729, 731, 762, 773, 781, 782
De Rijk, L. M., 47
decent logic, 1063, 1078
decidability, 500, 730, 740, 769, 775
  and philosophical interest, 769
  of **L** (and fragments), 778
  of **T**, **E**, **R** (and fragments), 736, 768, 769, 781
  of typability, 749
decision problem, 206, 215, 411, 415, 500, 501, 514, 515
Dedekind, J. W. R., 51, 412, 413, 448, 510, 527, 528, 531, 549, 550
deduction, 1067
deduction theorem, 727, 735
  for $\mathbf{R}_{\rightarrow}$, 735
  for $\mathbf{T}_{\rightarrow}$, 735
deductive
  closure, 1079, 1100
  inference., 205
  invalidity, 143, 152
  logical invalidity, 152
  theory, 53, 55, 56
  validity, 133
default, 1111, 1117
  logic, 46
  position, 1121
definite descriptions, 149, 972
definition, 41, 42, 46, 52, 60, 167
defuzzification method, 875
degree
  introduction rule, 712
  of truth, 166
  of confidence, 884, 889
  of unsolvability, 516
dense open orderings, 591
deontic logic, 898
dependence artificial, 1138
derivations and proofs, 554
derived rule, 752, 756
Descartes, R., 59, 901, 943
descriptions, 817, 872, 1076
designated element, 951
designated truth degrees, 679
determinable, 161, 162, 176
Detlefsen, M., 414, 420, 433–435
Deutsch, H., 1070
deviant logic, 165, 166
Devine, P. E., 1061
Dewey, J., 851
diagonal constructions, 143
diagram, 123
  box-arrow, 110
  language, 827
diagramming, 133, 135
Dialectica interpretation, 437
dialectical tier, 118
dialetheic 326, 367–376, 385, 404, 793, 873, 874, 1100, 1115
  approach, 900
  logic, 1069, 1079
dialogue, persuasion
  permissive, 124
  restrictive, 124

Díez, G. F., 557–558
dilemma, 139–141
3-dimensionalism, 979
4-dimensionalism, 979
Diodorus Cronus, 943
Dirac, P. A. M., 792
discourse representation theory, 82
discretion, 159
discursive logic, 792
discussive connectives, 846, 857
discussive logic, 792, 796, 844, 851, 855, 858
  and empirical truth, 851
  first-order calculus, 848
disjunction property, 757
disjunctive syllogism, 213, *see* γ
Divers, J., 930
Doležel, L., 1080
domain of existence, 961
double prevention, 1141, 1142
double-aspectness, 1108
Dowden, B., 367
downward Löwenheim-Skolem theorem, 592, 593
Dowty, David, 957
dtruth, *see* truth
Du Bois Reymond, 842
dual gaggle semantics, 757–761, *see* relational semantics
Dubikajtis, L., 844, 871
Dugundji, J., 220
Dummett, M., 384, 452, 458, 525, 535, 544, 548, 554–557, 559, 564, 567, 569, 913
Dunn, J. M., 724, 729, 731, 736–738, 741, 743, 751, 755, 758, 763, 764, 772–774, 777, 781–783, 785
Duns Scotus law, 794
Dylan, B., 550
dynamic
  epistemic logic, 90
  logic, 68
  predicate logic, 82
  semantics, 82

E model structure, 206, 210, 211, 220, 224
E (logic of entailment), 724, 735, 779–781
E-valid, 212
e-logic
  broadly free, 1077
E-postulates, 210, 215
early preemption, 1131, 1132
Earman, J., 577–578
Ebecken, N. F. F., 875
$\mathcal{E}$-connections, 973
ECP, *see* Exhaustive Characterization Project
Eddington, 1064
Edgington, D., 166
Ehrenfeucht, A., 595
Einstein, A., 165, 955
Elamrani-Jamal, A., 44
elementary extension, 592
elementary substructure, 592
Elga, A., 1130
elimination argument, EA, 213
elimination of quantifiers, 519
eliminativist, 985
ellipsis, 1069
embedding, 121
Emmy, the robot, 874
empirical model, 1019
empirical truth, 850
empiricists, 479
Ems, 210–218
endurantism, 979
Engel, S. M., 151
English predicate, 1013
entailment, 143, 205, 208, 211, 213, 217, 723, 729–782, *see* implication, 1065, 1067
Epimenides, 136, 137
epiphenomenal by-product, 1131
epistemic
  inconsistency, 875

  indeterminacy, 1111, 1112
  integration, 1105
  logic, 89, 165
  subworld, 1112
  epistemicist, 155, 165, 170
epistemological case against actual infinity, 547
epistemology, 50, 52
ϵ-calculus, 417, 418, 1077, 1097
ϵ-substitution method, 417, 430, 436
ϵ_0, 419, 435, 436
epsilon term, 1077
equality, 815
errors, as an argument against constructivism, 551, 560
Essenin-Volpin, A. S., 815
Etchemendy, J., 137, 389–391, 506, 918
Euclid, 46, 164
  geometry, 132, 831
Euler, 133
  diagrams, 58
European Union, 160
evaluated formulas, 710
Everett, A., 1082
Everett, H., 954
evidence in constructive logic, 556–558
Ewald, W., 527
ex falso quodlibet, 1080
examination of argument T, 1012–1021
excluded middle, 139, 140, 563
exhaustive characterization project, 328, 330, 331, 334, 339, 343, 344, 347, 348, 359, 367, 369, 371, 375, 376, 383, 387, 389, 391, 396, 402
existence, 1076, 1090, 1092
existence predicate, 1076
existence-neutral, 1061
existential quantifier, 133, 163
$\exists$, constructive interpretation, 552, 554–555
expansion, 589
expert systems, 865, 890
explanation, 77
explicit mathematics, 437
expulsion of contradictions, 1102
extension, 463, 916, 918–920, 922–932, 934–940
extensionality, 150, 151
extensional logic, 918
extensional semantics, 921
extranuclear properties, 1090

façon de parler, 1120
fallacy, 42, 47, 49, 125, 145
Falsehood Principle, 962
falsity free fragments, 708
familiar inference, 1100
family resemblance, 167, 170
faultless disgreement, 169
Faust, 1113
fde, 213
fde, *see* first-degree entailment
feasibility, 526, 551, 558–564
feasible arithmetic, 561–563
Feferman, S., 51, 340, 344, 345, 435–609
fictional
  creativity, 1096
  discourse, 1064, 1074
  entities, 1099
  object, 1090, 1100, 1110
  stipulationism, 1061
  world, 1097, 1098, 1114, 1120
fictionalism, 1104
fictive
  operator, 1093
  discourse, 1074
  modality, 1072, 1114
  operator, 1072, 1073
  qualifier, 1100
  Fidel, M., 825
  Field, H., 342, 343, 346–361, 363, 367, 369, 372–374, 377, 379, 381, 392, 397, 400, 420, 525, 1104, 1136

Fields medal, 892
filter, 758, 765, 767, 778, 784
Fine, K., 329, 367, 743, 758, 945, 948, 962
finite
  annotation property, 862
  characteristic matrix, 220
  matrices, 217
  model property, 757, 769, 778597
  model theory, 610
  occurrence property, 602
  trivializable, 809
finitism, 416, 420–427, 525, 1110
finitistic, 209, 218
Finocchiaro, M., 123
(first) incompleteness theorem, 504
first degree entailment, 213, 729–731, 754
  in **T**, **E**, **R**, **L**, 774
  perfect interpolation in, 731
first incompleteness theorem, 506
first-order
  definable, 517
  languages, 56
  like logics, 608
  logic, 46, 58, 498–500, 519, 589
Fitch Paradox, 92
Fitch-style calculus for $\mathbf{T}_{\rightarrow}$, 746–748
Fitting, M., 339, 859, 943, 959, 972, 976, 981
flea algebras, 709
Fletcher, P., 525, 526, 558
Flum, J., 607
Fodor, J. A., 560
'for any' versus 'for all', 553, 557
Forbes, G., 967
formal, 132, 136, 471, 1064
  dialectic, 124
  language, 49, 137
  logic, 131, 133–136, 206
  logical properties, 133
  provability, 206
  reasoning, 206
  semantics, 205–207, 209, 210, 213, 218
  -informal logic, 134
foundational reduction, 438–439
four-valued semantics, 729, *see* semantics of relevance logics
Fourier series, 842
Fraassen, B. van, 1077
Fraenkel, A. A., 449, 516, 517, 587
fragility strategy, 1133
free logic, 961, 1076, 1097
  in the broad sense, 1076
  in the narrow sense, 1076
Frege, G., 41, 43, 48–52, 56, 60, 134, 411, 413, 414, 421, 429, 433, 441, 449, 497, 498, 510, 526, 527, 531, 536, 537, 551, 725, 913–920, 923–926, 934, 936, 940, 941, 958, 1062, 1067, 1093, 1107, 1108
  Frege's Basic Law V, 531
  Frege's context principle, 531
  Frege's maneuver, 915, 916, 920, 923
Fregian, 1107
French, S., 795, 857
Friedman, H., 439, 607
Friedman, K. S., 572
Fuhrken, G., 599
Fuller, G., 1077
fully absolute, 609
function, 914–916, 918–924, 926, 930–932, 934–937, 939
function symbols, 589
functional interpretation, 437
fundamental logic, 206
fundamental theorem of the integral calculus, 1003
fusions of modal logics, 972
future contingent propositions, 166
fuzzy logic, 166, 167, 865, 873, 875

Gödel, K., 44, 51, 55, 60, 412, 418, 419, 431–435, 437, 439–441, 449, 475, 498, 502–506, 510,

512, 513, 515–518, 793
    Gödel logic, 683, 687
Gabbay, D. M., 1112, 1117, 1121
Gabriel, G., 1061, 1082, 1099
Gaifman, H., 384, 404
Gale, R. M., 1061
Gallin, D., 932, 934, 935, 959
game, 68, 84, 92, 93, 1081, 1083, 1088
γ (rule), 773
    admissible in **T**, **E**, **R**, 773
Gamut, L. T. F., 926, 930, 932
gap, 694
gap-1 theorem, 601
gap-2 theorem, 601
Gardner, M., 158
Garth Mountain, 160
gates, 880
general frame, 950
    compact, 950
    descriptive, 950
    differentiated, 950
    logic system, 602
    recursive functions, 513–516
    tight, 950
generalised
continuum hypothesis, 516
    coherence frame, 974
    Kripke-frame, 965
quantifiers, 67
generalists, 164
genericity, 1111, 1112
Genesis, 1103
Gentzen, G., 55, 411, 419, 432, 435–437, 441, 505, 508, 509, 519
    Gentzen's consistency proof, 514
genuine logic, 207, 209
genus/differentia distinction, 167
geometry, 46
George, A., 551, 559
George, R., 122
Gerla, G., 710, 712
Giambrone, S., 738, 740, 755, 768
Giaquinto, M., 429, 430
Gilbert, M. A., 105
Girard, J.-Y., 724, 779
Giuntini. R., 840
Glanzberg, M., 386–389, 399
glut, 694
God, 146, 147, 149–152
Gödel, K., 542, 564, 726, 783
    Gödel's 'Dialectica' interpretation, 555
    Gödel's completeness theorem, 497
Goethe, 1113
Goguen, J., 166
Goldbach's conjecture, 555, 565, 573, 574
Goldblatt, R., 929, 955, 964
golden mountain, 1090, 1098
Goldfarb, W. D., 936
Goldman, A. I., 124
Gonseth, F.
    on the nature of logic, 795
good argument
    dialectical criteria for, 116–117
Good Samaritan paradox, 898
Goodman, N. D., 557–558, 1128
Grana, N., 872
Granger, G. G., 796
Greenough, P., 403
Grice, P., 1069
Griffin, N., 1074, 1097
Grim, P., 328
Groarke, L., 58, 104
Grohe, M., 611
ground literal, 876
Grünbaum, A., 567, 568, 570–572
Grundgesetzte der Arithmetik, 510
Grundlagen der Mathematik, 505
Grundzüge der theoretischen Logik, 500
Guaspari, D., 435
Gupta, A., 339, 392, 395, 396, 402, 405

H (intuitionistic logic), 728, 737, 740, 757
    implicational fragment of, 749, 754

    in **L**, 778, 779
    in $\mathbf{R}_+$, 783
    philosophical import of, 783
Haack, S., 159
haecceitism
    D-haecceitism, 980
    first-order haecceitism, 979
    second-order haecceitism, 980
Hájek, P., 702, 709
Halbach, V., 403, 405
Hale, B., 441, 465
Hall, N., 1127, 1134, 1139, 1142
Hallett, M., 414, 429, 430, 534, 541, 542, 544, 545, 551
Hamblin, C. L., 116, 124–125
Hanf number, 603
Hanf, W. P., 604
Hanle algebra, 846
Hanley, R., 1087, 1102, 1103
Harman, G., 380
Harrington, L., 440
Harrison, C., 571
Hartshorne, C., 147, 149, 150
Hausdorff, F., 516
Hazen, A., 968
Heck, R., 441
Heintz, J., 1079
Heisenberg, W., 897
    uncertainty principle, 896
Hellman, G. H., 525, 565
Hellman, G., 748
Henkin, L., 497, 507, 519, 923, 960
    Henkin semantics, 923
    Henkin set, 827
Henry, D. P., 149
Herbrand, J., 419, 513
    Herbrand base, 867
    Herbrand interpretations, 867, 868
hereditary right maximal terms, 750
    simple types of, 750
heredity lemma, 744, 759, 764
Herzberger, H., 352, 392
heuristic, 206, 210, 220, 225
Heyting, A., 546, 548, 550, 552, 553, 944
    Heyting algebras, 687
Hick, J., 147
Hilbert, D., 44, 53, 54, 57, 207, 208, 433, 434, 441–1104
    Hilbert spaces, 524
    Hilbert's program 411–441
        and instrumentalism, 428–435
        history of, 412–419
    Hilbert's first problem, 517
    Hilbert's second problem, 516
Hilbert–Bernays' theorem, 814
Hilbertian, 1107
Hindley, J. R., 723, 749, 776
Hintikka, J., 928, 939, 987, 1088
history, 957
history-constitutive, 1093, 1096
Hitchcock, C., 1138
Hodes, H. T., 1104
Hofweber, T., 439
homomorphism, 914
hoops, 708
Hopkins, J., 147
horizon problem, 525–526, 546, 563, 564
Horsten, L., 403, 405
Hovweber, T., 1082
Howell, R., 1061, 1080, 1091, 1092, 1099
$\mathbf{HRM}_z$-terms, *see* hereditary right maximal terms
Hrushovski, E., 49
Hughes, G. E., 926
Hume, D., 479, 1130
    Hume's Principle, 465
Huntington, E. V., 497, 498
Hurford, J. R., 51
Husserl, E., 60, 414
Hyde, D., 329
hyper-literal, 878, 886, 887
hyper-reals, 841
hyphenation, 1104, 1116
hyphenation-blindness, 1105
hyphenated truth, 1105

hypothetical object, 1079

I (combinator), 733, 736, 738, 749, 750, 751, 780
  simple type of, 723, 733, 751, 776
Ibn Sīnā, 41
ideal observer, 1113
idealisation, 559–561
idealism, 1110
idealist
  instrumentalism, 1110
  stipulation, 1120
identity of indiscernibles, 795
identity sign, 133
Ignjatović, A., 434, 435, 1110
ill-behaved formulas, 799
illative
  conclusion indicator, 106
  core 118
  premise indicator, 106
imaginary logic, 792, 794
immigrant object, 1090, 1092, 1097
impossibilia, 1098
imperfect consistency, 212, 214, 215
imperfective paradox, 957
implication, 723, 729–784, *see* entailment
$\rightarrow$ (implication), 723, 729
$\Rightarrow$, constructive interpretation, 554–558
impossibilia, 1065, 1094
impossibility, 1099
impredicative, 416, 488
impredicativity of constructive proof, 557
inclusion conditiuon, 211
incompleteness, 215, 216, 1099, 1103
  theorem, 412, 418–419, 431–435, 440, 441, 503, 505
    first, 418, 433, 434
    second, 418, 431, 432
inconsistent, 141, 1104
  objects, 901, 902
  reasoning, 901
  theory, 791, 797
indefinitely extensible concepts, 544
indeterminacy, 1103
Indian logic, 42
indispensability argument, 438
individual, 959, 976
  concept, 945, 959, 973
individuation, 1097
induction, 463
  contentual, 425
  Poincaré on, 414
  transfinite, 425, 435, 436
inertia worlds, 957
inexactitude, 157
inference, 41, 46, 105, 117–118, 1064, 1066, 1067, 1068, 1119
inference rule, 163
inferential closure, 1112
infinitary logic, 54, 593
infinite, 488
  sums, potential and actual views of, 523, 567, 576
  quantifier, 592
  variable, 841
infinitesimal, 524, 525, 841
  variable, 840
informal, 132
informal logic, 58, 131, 133–136
Ingarden, R., 1091
ingenuity, 169
inherent ambiguity, 871
inheritance networks, 859
inner model, 517
instrumental, 1110, 1120
  necessity, 1063, 1065
  stipulationism, 1120
instrumentalism
  and Hilbert's program, 431
  and Hilbert's program, 428–435
intensionality, 150
intensional 151
  logic, 925, 930–932, 934–936, 938
  modal fallacy, 146
interaction, 68

internal set, 950
interpolation (Craig), 731, 770
  perfect, 731
intrinsicness, 1141, 1142
introspection, 56, 60
intuition, 50, 56, 478, 411, 415, 420, 430, 525, 558
  and finitism, 426
  Kantian, 423
intuitionistic logic, 791, 794, 814, 897
  Brouwer-Heyting system, 794
  contained in $\mathcal{C}_1$, 803
  minimal calculus, 814
intuitionistic propositional logic, 797
intuitive axiom, 1097
intuitive axioms for fiction, 1068
invalidity, 134
Inwagen, P. van, 1070
irregular verbs, 160
irregular words, 164
iterative conception of sets, 533, 541–545

Jäsche, G. B., 42
Jaśkowski, S., 792, 793, 844
  Jaśkowski logic, 844, 871
Jacquette, D., 149, 331, 1074, 1078, 1079, 1087, 1092, 1097, 1099–1101
James, H., 1088
James, W., 851
Janssen, T. M. V., 913, 916
Jennings, R., 1079
Jensen, R. B., 601
Jeroslow, R., 435
Johnson, R. H., 101, 118, 122, 126
Johnson, W. E., 161, 167
Johnstone, M., 979
Johnstone, P. T., 567
join, 218, 219, 221
justifying formal logic, 207

K (combinator), 733, 736, 749, 779
  simple type of, 723, 727, 735, 749, 777

Künne, W., 403
Kaestner, C. A. A., 876
Kant, I., 42, 43, 51, 57, 58, 134, 146, 149, 151, 152, 164, 478, 1110
  Kant's 100 gold Thalers, 151
Kaplan, D., 327, 980, 1061, 1068, 1099
Keisler, H. J., 595
Kelley-Morse set theory, 830
Kepler, J., 157
Kifer, M., 858, 859
Kilwardby, R., 42, 44, 45
King, L. R., 1075, 1087
Kitcher, P., 421, 423, 429
Kleene, S. C., 498, 511–513, 515, 519, 555, 556, 682, 689, 802
Kleer, J. de, 1078
Kneale, M., 943
Kneale, W., 943
knowledge
  acquaintance, 1076
  description, 1076
  ordering, 696
Kock, A., 528
Koetsier, T., 572
Kolmogorov, A. N., 552, 554
Koons, R., 384, 386, 404
Kotas,J., 871
Krabbe, E. C. W., 125
Kracht, M., 943
Krause, D., 876
Kreisel, G., 48, 426–609
Kremer, M., 340
Kremer, P., 340
Kripke, S., 138, 143, 326, 330–335, 337–345, 347, 348, 350–352, 354, 356, 357, 359, 360, 363, 367, 369, 371, 373, 383, 392, 736, 763, 927, 928, 937, 945, 960, 961, 964, 1061, 1068, 1070, 1088, 1098, 1107
  Kripke model, 856
  Kripke semantics, 847, 873
  Kripke structure, 846, 847, 851
  Kripke-frame, 946

  Kripkean, 141
  pointed Kripke-model, 946
Krishnaprasad, T., 859
Kronecker, L., 412–413, 531, 563
Kutz, O., 943

L (linear logic), 775–779
l'Hospital principle, 840, 842, 843
Löwenheim, L., 587
  Löwenheim number, 603
  Löwenheim, L., 499, 598
  Löwenheim–Skolem Theorem, 498–499, 519, 587, 590, 616, 645–647
López-Escobar, E. G. K., 605, 607
labels, 1116, 1117
Lake, J., 539, 540
Lakoff, G., 170
Lamarque, P., 1070
λ
  definability, 512–515
  abstraction, 933
  calculus, 511, 723, 737, 749–751, 779, *see* combinatory logic
  conversion, 923
  operator, 133
Lambert, K., 961
Landman, F. W., 1078
language, 134, 136, 140, 146
late preemption, 1132–1134, 1137, 1139, 1140, 1141
lattice, 218–220
Law of Excluded Middle (LEM), 327, 330, 331, 337, 342, 345, 347, 349–351, 356, 357, 359, 360, 367–369, 375, 380, 384, 385, 1067
law of noncontradiction, 1067, 1094, 1095, 1100
Lawvere, W., 50
Lear, J., 523, 567, 1091
learning, 87
legal obligations, 898
Leibniz, G. W., 41, 45, 46, 48, 49, 58, 529, 795, 944
  Leibniz Law, 962
  Modal Leibniz Law, 963
  strong Leibniz Law, 962
  weak Leibniz Law, 963
Lejewski, C., 1076
LEM, *see* Law of Excluded Middle
Levy, A., 597, 609
Lewis translation, 967
Lewis, C. I., 369, 373, 926, 944, 952
Lewis, D., 479, 529, 929, 937, 939, 945, 948, 956, 962, 967, 979, 980, 1061, 1070, 1079, 1086–1088, 1090, 1094, 1099, 1101, 1103, 1128, 1130, 1132, 1133, 1139, 1141, 1142
  Lewisean truth I, 1087
  Lewisean truth II, 1087
  Lewisian counterpart frames, 977
liar, 143, 145
  dilemma inference, 143
  metaparadox, 140, 141
  paradox, 136–145
  sentence, 138–146
  sentence dilemma, 139, 140
  sentences, 137
limitation of size, as a basis for sets, 537–541
Lindenbaum, A., 507
  Lindenbaum algebra, 686, 949
Lindström, P., 451, 499, 588, 607
  Lindström's theorem, 607
linear logic, 954
linguistic ersatzism, 949
literacy primacy thesis, 1069, 1104, 1117
literals
  complementary, 881
Lobachewski, N., 792
logic, 42, 52, 602
  C, 964
  PFL, 963
  QC, 964

  - **QK**, 960
  - **Q**L, 961
  - **S3**, 944
  - **Grz**, 953
  - **Int**, 953
  - **N**, 953
  - **QS4**, 970
  - **S4.2**, 955
  - **S4**, 953, 956, 969
  - logic of $\mathcal{K}$, 947
  - algebraic, 414
  - of conditionals, 956
  - deontic, 898
  - of fiction, 1061, 1065–1067, 1069, 1074, 1078, 1099, 1100, 1112
  - first-order, 415
  - imaginary, 792
  - informal, 101
  - of Minkoski space-time, 955
  - modal, 898
  - nature of, 793
  - and physics, 892
  - pluralism, 902
  - pure, 51
  - possibilistic, 875
  - scientific acceptance, 857
  - of science, 857
  - of truth, 857
  - logic, universal, 51
  - of the world, 901
- logical
  - antinomy, 145
  - concepts, 43, 45, 51, 56, 57
  - consequence, 55, 56, 680
  - constants, 206
  - constants, constructive interpretation, 553–558
  - empiricism, 892
  - entailment, 208, 211, 217
  - fatalims, 166
  - form, 67, 133, 135, 142
  - matrix, 951
  - necessity, 964
  - pluralism, 902
  - positivism, 473
  - relations, 131
  - signs, 215
  - truth, 53, 964, 1065, 1067
  - valid, 679, 686
- logicism, 50, 415, 420
- logistic paradigm, 53–56
- Loparić, A., 825, 870
- Loux, M. J., 930
- Lozinskii, E, L., 859
- Luce, A. A., 57
- Ludwig, G., 893
- Łukasiewicz, J., 166, 680, 684, 687, 693, 792, 829, 844, 871, 943, 1078
  - Łukasiewicz algebras, 687
  - Łukasiewicz logic, 687, 684
- Luria, A. R., 59
- Lynch, M., 384

- M (combinator), 738, 741, 779
  - simple type of, 749
- $\mathbb{M}$ and M, 616
- Müller, A., 422
- Mackie, J. L., 1131
- Maddy, P., 527, 542–543, 1081, 1110
- make-believe, 1084–1087
- make-do, 1081, 1082, 1088, 1099, 1115
- Malcolm, N., 146
- Mally, 1092
- Malt'sev, A. I., 507, 519
- Mancosu, P., 414, 418, 423, 430
- Manin, Y. I., 842
- many worlds interpretation, 954, 986
- many-valued logic, 166
- Marconi, D., 870, 873
- Marcos, J., 872
- Marcus, G. F., 51
- Marcus, Ruth Barcan, 960
- Mares, E. D., 373, 724, 732, 781
- Marin, R. L., 333, 403
- Markov, A. A., 558
- Mars, 157
- Martin, D. A., 49

Martin, R. L., 137, 330, 333, 335, 338, 340, 403, 1061
Martin-Löf, P., 555, 556
Martins, A. T., 875
Master Argument, 943
material conditional, 145
⊃ (material implication), 726, 727, 728
  definition, 1071
  induction, 155, 210
  logic, 44, 46, 49, 131, 135, 1062, 1063, 1065–1067, 1080, 1099
  multiplicity, 142
  stipulationism, 1061
Matilal, B. K., 42
matrices, 879
  as trees, 877
  contradictory, 880
  valid, 880
matrix connection method, 876
matrix suitable for a system, 220
matrix tables, 221
Mattā ibn Yūnus, 44
mature, 1063, 1065, 1066
Maudlin, T., 340, 344
maximal account, 1067, 1075, 1079, 1080, 1090, 1100, 1111, 1113, 1115, 1119
maximality, 1111
Mayberry, J. P., 527, 534, 538–542, 549–550
McGee, V., 330, 339, 348, 349, 398
McGill, A. C., 147
McGill, N. J., 874
McKeon, M., 1108
McRobbie, M. A., 777, 778, 781, 782
meaning, 136, 459, 913–917, 924, 935, 939
medieval logicians, 41, 44, 47
meet, 218, 219, 221
Meheus, J., 1063, 1078, 1079
Meinong, A., 449, 1104, 1091
  Meinongian creativity, 1096
  Meinongian logic, 1078, 1090, 1097, 1101
  Meinongian object, 1090, 1094, 1095, 1100
  Meinongian ontology, 1103
  Meinongian world, 1099
Mekler, A., 595
Melia, J., 980
Mendelson, E., 918
mental language, 48, 60, 160
Meredith, C. A., 752
mereology, 165, 529, 569
metalanguage, 43, 45, 55, 58, 136, 142, 143, 207, 209, 213
metamathematics, 54, 416
metaparadox, 140, 145
metaphysical footprint, 1070, 1072, 1073, 1087, 1107, 1110, 1119, 1121
method
  of constants, 519
  of intersection, 1088
  of union, 1088, 1102
Meyer, R. K., 212, 345, 724, 754, 758, 763, 764, 773–775, 777, 778, 781–783, 1079
Michelson, A. A., 792
Michelson–Morley's interferometer, 792
migrancy, 1097
Mikenberg, I., 851
minimal adaptation, 1066, 1076–1079
Minkowski space-time, 955
Mints, G., 436
Miró-Quesada, F., 793
Mirimanoff, D., 542
misleading form thesis, 69
Mitchell, O. H., 526
mixed sentence, 1075
ML language, 883
modal, 146
  algbera, 949
  context, 150, 1073
  hyperdoctrine, 983
  individual, 979
  logic, 165, 166, 723, 743, 746,

        757, 768, 777, 780, 898, 926, 946, 1128
      KB, 763
      S4, **S4.2**, **S4.3**, 763
      classical, 946
    normal, 946
    necessity operator, 133
    operator, 1094, 1101, 1115
    operators, 946
    predicate logic, 958
      first-order, 959
      second-order, 959
    propositional logic, 946
modality, 946
    de dicto, 1072
    de re, 1072
modally fragile, 1133, 1134
model, 679, 710
    classes, 679
    structure, 210
    structure semantics, 220
    structures, 217
    theory, 41, 519
modes of transgression, 969
modist grammar, 48
modularity trigger, 1121
modus tollens, 147, 149
Moh Schaw-Kwei's paradox, 830
monad, 841
monadic fragment, 972
Monk, R., 144
monotonicity, 67
Montague, R., 930–935, 959
    Montague Grammar, 913
    Montague's Thesis, 74
Montague, R. M., 597
Mooij, J. J. A., 1082
Moore pragmatic paradox, 144
Moore, G. E., 144, 415, 456
    Moore sentence, 144, 146
    Moore's paradox, 144
Moore, G. H., 529
moral dilemmas, 898
moral obligations, 898
morality and law, 898
More, T., 47
Morgenstern, O., 1088
Morley, E. W., 792
Morley, M. D., 605, 607
Mortensen, C., 843, 871, 873
Mortimer, M., 598
most general common instance (*mgci*), 752, 753
Mostowski, A., 595, 599
    1 Mostowski Collapse, 620
MTL-algebras, 705
μ-recursiveness, 514, 515
Musicante, M. A., 876
MV-algebras, 687

N, the set of natural numbers, 523, 524, 548
naïve idealism, 1066
naïve set theoretical, 145
Nakamatsu, K., 874
narrow free e-logic, 1076
Nash equilibria, 89
native, 1090
native objects, 1097
natural
    deduction, 58, 508
    language, 134, 135
    language interpretations, 1013
    logic, 72
    numbers, 547, 548
    reasoning, 851
    speech, 1073
necessitation rule, 1098
necessity, 924, 926
    of identity, 960
negation, 729, 769, 773
-as-failure, 1117
complement, 219
    De Morgan, 729, 731, 762, 773, 782
    ortho-, 762, 784
    strong negation, 804
    the nature of, 794

negative existential, 1076, 1080
neighbourhood frame, 955
Nelson's inconsistent logic, 874
Nelson, L., 414, 423
Nelson, J. O., 147
Nelson, N. D., 953
neo-Fregean, 464
neural nets, 170
Newton, I., 157, 165, 527
    cosmology, 792
    second law, 850
Nietzsche, F., 156, 157
no-class theory, 467, 471
no-counterexample interpretation, 440
Nobre, E. F., 876
nominal definition, 1088
nominalism, 48, 478, 1119
nominalistic instrumentalism, 1110
nominalistic provocation, 1114
nominalization, 900, 1119
non-Aristotelian logics, 792
non-Euclidean geometry, 132, 792
non-individuals, 795
    in the quantum domain, 796
Non-Triviality Project, 328, 330, 334, 337, 339–344, 347, 348, 356, 367, 371, 375, 376, 381, 383, 386, 389, 391, 395, 402
nonbivalent, 1077
nondenoting terms, 1076
nonexistence, 1149, 51, 1069, 1076, 1090, 1092
nonmonotonic logic, 79, 1128
nonnormal world, 1098
nonnuclear proerties, 1092
nonstandard
    analysis, 44, 528
    arithmetic, 561
    many-valued logics, 137
    models, 503
nontrivial, 897
normal Ems, 213
normal form theorem, 514
normal negation, 219
Norman, J., 142
Norton, J. D., 577
$\neg$, constructive interpretation, 554
NTP, *see* Non-Triviality Project
nuclear property, 1090–1093, 1097, 1101
nuclear–nonnuclear, 1094
number, 52, 60, 162
Nyāya logic, 42

O-interpretation, 1072
O-intuition, 1070, 1071
O-model, 1072–1074, 1082, 1083
Oakshott, M., 1082
object, 945, 973, 976
    language, 43, 142, 143, 207
    metalanguage, 138
    native, 1092
    variables, 958
objecthood, 1078
objectual quantification, 1108
Ockhamist, 957
Ogden, C. K., 142
Olsen, S., 1070
$\Omega(x)$, 617
$\Omega_E(x)$, 626
$\Omega_{\mathbb{M}}(x)$, 626
ω-consistency, 504, 506
ω_1, the first uncountable ordinal number, 538
ω-rule, 432, 434
omission, 1135
ontically incomplete, 1091
ontological argument, 149, 150, 152
ontological case against actual infinity, 547
open textured, 169
operational semantics, 743–746, 777, *see* semi-lattice semantics, semantics of relevance logics
$\vee$, constructive interpretation, 553–555
order-type, 530
ordinal analysis, 432, 436
ordinal sum, 701

ordinary language, 135
ordo cognescendi, 1112
ordo essendi, 1112
origins, 211
overcomplete theory, 856
Oversteegen, J. J., 1082
Owen, G. E. L., 567, 569

Péter, R., 511
para-analyser, 874
paraclassic logic, 793, 851, 894, 895, 897, 898
  paraclassic deduction, 894
paraclassic theory, 898
paracomplete, 325, 326, 329–331, 334–339, 342–351, 356, 357, 359–361, 364, 366–370, 372, 375, 378–381, 398, 401, 404, 874, 888
paraconsistent, 143, 325, 326, 328–330, 343, 353, 357, 361, 366–368, 370, 374, 375, 378–380, 382, 398, 400, 401, 404
  dC-systems, 872
  $\mathcal{C}$-systems, 796
  algebraization, 871
  agnostic interpretation, 902
  applications, 794, 865, 874
  Aristotle's syllogistic, 829
  Boolean algebra, 839
  complementarity, 894
  dialectics, 873
  calculi $\mathcal{C}_\omega^=$, 815
  calculi of descriptions, 817
  calculi $\mathcal{C}_n$, 797
  consequence, 1079
  decidability, 870
  deontic, 898
  executing queries, 870
  field of knowledge, 902
  forerunners, 792
  fuzzy systems, 875
  hierarchy of calculi, 807
  higher-order systems, 872
  in computer sciences, 858
  in Informatics, 875
  inductive, 884
  its role in science, 792
  logic, 142, 791, 1079
  logic of partial truth, 851
    as 'supplement' to classical logic, 795
  mathematics, 796, 840
  Mathematical Reviews, 793
  modal, 829
  model for the differential and integral calculus, 842
  model theory, 828
  negation in, 798
  origins of the name, 792, 793
  paraclassic logic, 898
  philosophical significance, 898
  predicates, 829
  program, 881
  'pure' logic, 794
  quantification, 811
  quantum logic, 840, 892
  reductio ad absurdum principle, 800
  'rival' to classical logic, 795
  robot, 874
  set theory, 791, 830, 831, 901
  syllogistic, 828
  taxonomy, 872
  theory of v.b.t.o., 872
  three valued systems, 871
  valuation, 843
  World Congresses on Paraconsistency, 793, 902
paradox, 137
  Cantor's, 726, 727
  geometric, 567–569
  kinematic, 569–575
  of collision, 576–578
  of conserved quantities, 575–576
  positive and negative, 723, 727
  semantic, 537
  set-theoretic, 531–532

  - Zeno's, 567–571, 576

PARALOG, 874

paralogic, 897, 898

paranormal, 350, 361–366, 369, 381

parasitical, 1114, 1120

Parikh, R., 561–563

Paris, J., 440

Parmenides, 158

Parry, W. T., 874

Parsons, C., 387, 388, 423, 426, 544, 1061, 1067

Parsons, T., 372, 962, 1074, 1075, 1078, 1080, 1090–1092, 1097

Partee, B., 932

partial
  - object, 1078
  - preemption, 1139
  - recursive function, 515
  - truth, 851, 857, 901, 902

particularists, 164

particulars, 479

Pascal, B., 41

Pasch, M., 497

path through a matrix, 880

Patterson, D., 382

Paul, L. A., 1127, 1134, 1139

Pauli's exclusion principle, 897

Pavel, T., 1061, 1080, 1099

Pavelka, J., 710
  - Pavelka-style approach, 710

PBM-model, 1072

PC, 208

Peña, L., 871, 873

Peano, G., 41, 49, 449, 497, 498, 502, 510
  - Peano arithmetic, 419, 420, 432, 435, 437–440, 503, 505, 507, 508

Pearl, J., 1138

Peckhaus, V., 414

Peirce, C. S., 45, 46, 49, 134, 497, 499, 526, 528, 567, 792, 851
  - Peirce's law, 723, 736, 806, 832
  - Peircean, 958

Pelletier, F. J., 916, 1112

Penrose, R., 158

Pequeno, T., 875

perdurantism, 979

Peressini, A., 565

Pérez Laraudogoitia, J., 576–578

perfect recall, 91

performative utterance, 1069

persistence, 979

Peter of Ailly, 60

Phillips, J. F., 1087

philosophical
  - analysis, 1071, 1088
  - logic, 1062, 1063, 1065
  - semantics, 145

philosophy of language, 1062

philosophy of mathematics, 1061, 1104

physical infinity, 564–579
  - in relation to mathematical infinity, 525, 564–566

physical theories, 849
  - structures for, 849

Picard and Gousart's treatise, 840

picture theory of meaning, 142

Pigozzi, G., 1121

Pinker, S., 164, 165

Pinter, C., 871

Pinto, R. C., 117

Placek, T., 553

Plantinga, A., 147, 151, 1061, 1068, 1099

Plato, 156, 158, 167, 1086

Platonism, 485, 525, 527
  - plenitudinous, 531

Plummer, C., 1084, 1085, 1108

pluralism, 1063, 1104

PMB-model, 1070, 1074, 1082, 1088, 1110

Pohlers, W., 436

Poincaré, J. H., 414, 531, 537, 540, 550

pointless topology, 567

points at infinity, 528

polarity of a literal, 876

Pollard, S., 539
Pollock, J., 884
poly-counterpart frame, 970
polynomial-time computability, 561
Poncelet's imaginary geometry, 831
Popper, K. R., 792
Porter, C., 1064
positive free logic, 963
positive implicational logic, 723, 748, 749
  R$_{\to}$'s relation to, 723
possibilistic logic, 875
possibility, 925–927, 929, 944, 952, 1092
possible object, 1079
possible worlds, 743, 757, 785, 925, 927–930, 1099, 1114, 1128
  possible worlds semantics, *see* relational semantics, dual gaggle semantics
Post, E., 498, 501, 514, 515, 680
  Post algebras, 688
  Post completeness, 725, 727
  Post logic, 685, 688
postulates, 217
  for model structures, 210
potential infinity, 523–525, 546–564
  versus actual infinity, 548–552
¶(S), the power set of S, 532
power set axiom, 532, 540–541, 543
Pozsgay, L., 544
pragmatic, 47, 49, 136, 152
  theory, 856
  truth, 844, 851
  valid formulas, 854
  validity
    extended to higher-order (modal) languages, 855
    the logic of, 855
precisification, 160, 163
predicate, 146
  complementation, 1092, 1103
  constants, 958
  logic, 207, 214, 215
  variables, 959
  -quantificational calculus, 134
predication, 452
predicativity, 537, 540
preemption, 1127, 1132
presheaf semantics, 969
presumptive closure, 1111
pretense, 1070, 1082–1084, 1086–1088
Priest, G., 142, 143, 149, 151, 360, 367–373, 375, 376, 398, 404, 873, 899, 900, 1079, 1087, 1098
primitive recursion, 423, 426
  and finitism, 425–426
  arithmetic, 425, 427, 435, 437, 438, 439, 1110
primitivist, 985
principal types, 723, 733, 735, 736, 749, 752, 754, 776, 779, *see* simple types
principle of
  bivalence, 675
  compositionality, 675
  explosion, 794, 872
  noncontradiction, 792, 798
Prior, A., 166, 948, 957
probability logic, 46
process logic, 56
procrusteanism, 1081
product algebras, 687
product logic, 684, 687
production, 1142
products of modal logics, 972
projective geometry, 770–771
PROLOG, 874, 882
proof, 476
  as normal natural-deduction derivation, 555
  canonical, 555
  constructive, 552–558
$\vdash$, the constructive proof relation, 552
  decidability of, 555–557
proof search, 41
proof theory, 41

  - complexity of proofs, 440
  - consequence, 55
  - ordinal analysis, 436
  - proof theoretic reduction, 438
  - reverse mathematics, 439
  - speed up, 435
- proofs, in axiomatic calculus
  - in $\mathbf{T}_{\rightarrow}$, 733–735
  - traversing vs ordinary, 733–734
- prop, 1083, 1085. 1088
- proper classes, 517
- propositional
  - attitude, 1082, 1083
  - closure, 1112
  - constants, 946
  - logic, 501
  - matrices, 876
- prototypes, 170
- provability degree, 711
- pseudo-questions, 474
- Pseudo-Scotus, 794
- $\Psi(f, x)$, 634, 637
- psychologism, 60, 913, 915
- Pudlák, P., 440, 562
- pure instrumentalism, 1110
- pure logic, 793
- Purtill, R. L., 147
- Putnam, H., 438
- *Px*, 210–212, 217
- Pylyshyn, Z. W., 560
- Pythagoreans, 158

- quadratic polynomial equations, 135
- quantification, 1061, 1114
  - prime formulas, 812
  - substitutional, 1111
  - theory, 497
- quantifier, 524, 528, 1073
  - actualist, 962
  - bounded and unbounded, 538
  - infinite disjunctions and conjunctions, 526, 565
  - interpreting functions, 677
  - possibilist, 962
  - rank, 589
- quantum logic, 794, 795, 952
  - the logic of quantum theory, 893
- quasi
  - -concrete, 423
  - -dependence, 1139–1141
  - -matrix, 823
  - -truth, 901
- question-begging, 1102
- Quine, W. V. O., 52–55, 59, 369, 438, 478, 537, 900, 926, 936–938, 986, 1063, 1064, 1076, 1081, 1114
  - Quine's NF, 830, 832
  - Quine's ontological commitment, 902
- quota model, 168

- **R**, the set of real numbers, 524, 528
- **R** (logic of relevant implication), 723, 735, 760, 769, 781–785
- Rüstow, A., 137
- *Rabc*, 210, 211, 215, 220, 224
- Radford, C., 1085
- Raggio, A., 871
- Ramachandran, M., 967
- ramified theory of types, 488
- Ramsey, F. P., 396, 441, 449, 611
- Ramus, P., 41
- random, 165, 169
- Rapaport, W. J., 1092
- Rathbone, B., 1083
- Rathjen, M., 436
- Read, S., 149
- realisability, 555
- realism, 526, 1071, 1072
- realist stance, 1105
- reasoning
  - defeasible, 874, 884
  - fuzzy, 886
  - nonmonotonic, 874, 884
  - styles, 67
- received theory, 1105
- Reck, E., 414

reductio, 149, 151, 208, 212–215
recursion theory, 498, 510,515, 516, 519
recursively enumerable set, 514
reducibility for constructive proof, 557–558
reduct, 589
reductionist, 985
reference, 136, 479, 1064, 1067, 1068, 1107, 1111, 1119
referent, 1107, 1111
Reflection Theorem, 597
regularity theory of causation, 1131
reification, 1071, 1073
Reinhardt, W. N., 330, 340
relation symbols, 589
relational semantics, 757–782, *see* semi-lattice semantics, semantics of relevance logics
  gaggle kind, 757, 758, 761, 763
  Kripke-style, 757, 763, 778
  philosophical interpretation of, 763
  Routley–Meyer-style, 758, 763–768, 775, 777, 781, 782
relations, 133
relative computability, 515
relativization, 589
relevance logic, 147, 870, 873, 954, 1064 *see* **B**, **T**, **E**, **L**, **R**, completeness, semantics (four-valued, semi-lattice, operational, dual gaggle, relational), classical logic
  modalities in, 773, 775, 777, 780
relevant
  arithmetic, 774
  entailment, 1063
  predication, 774
replacement axiom, 517, 532, 541, 543, 544
representation, choice of, 52
Rescher, N., 367
residuation, 699, 737, 742, 743, 760–762, 783, 784
Resnik, M., 414, 429, 1104
Restall, G., 346, 347, 364, 723, 724
restricted functional calculus, 500
reverse mathemtaics, 439
rhetoric of argumentation, 131
rhetorical fallacy, 134
Richardson, H., 147
Richman, F., 548, 554, 564
Riemann FOIL frames, 976
Rijke, M. de, 940
ring species, 161
R-mingle (semi-relevant logic), 729, 758
Robinson, A., 44, 507, 519, 528
robotics, 890
Romantics, 156, 157
Rosch, E., 170
Rosen, G., 962
Rosser provability, 434
Rosser, J. B., 498, 506, 511, 512, 514
round square, 1090
Routley, R., *see* Sylvan, R.
Routley, R., 142, 212, 346, 367, 373, 871, 873, 1061, 1070, 1072, 1075, 1077, 1079, 1080, 1093, 1100
Routley–Meyer-style semantics, 758, 763, 768, 775, 777, 781, 782
Russell, B., 46, 50, 145, 156, 157, 170, 208, 411, 413–415, 449, 498, 510, 525, 527, 532, 536–538, 540, 551, 557, 566, 573, 817, 833, 972, 1071, 1076, 1078, 1088, 1105
  Russell's antinomy, 497
  Russell relations, 838
  Russell set, 794, 796, 836, 900, 902
  Russell's paradox, 145, 397, 413, 451, 510
    semantical properties, 397
    set-theoretic, 396
Rxyz, 211
Ryckman, T., 430

S (combinator), 737, 749, 750
  simple type of, 736, 749, 779
S′ (combinator), 737, 748, 750
  simple type of, 736, 779
S-implication, 700
S-model, 1073, 1082, 1088, 1090, 1103, 1110, 1116, 1118–1120
safety-critical systems, 58
Salmon, W. C., 567, 573
salve veritate, 150
Sanford, D., 166
satisfiability problem, 597
satisfiable, 590
saving the appearances, 851
Sazonov, Y. V., 561, 562
Schütte, K., 436
Schaffer, J., 1133, 1135, 1142
Schantz, R., 403
Schlick, M., 474
Schopenhauer, F., 42
Schotch, P., 1061, 1079
Schröder, E., 414, 497
Schrödinger, E., 795
Schufreider, G., 147
Schurz, G., 964, 974
Schweber, S. S., 578
science, 132
Scott, D. S., 555
Scott–Montague frames, 956
Scruton, R., 1070
sea-battle, 166
Searle, J. R., 103, 1069, 1070, 1082, 1083, 1086
  Searle's Principle, 1083
second clauses, 556–558
second incompleteness theorem, 505, 508
second-order logic, 55, 502, 503, 592
self-contradictory, 144
self-countering threat, 1136, 1137
self-identity, 1092
Sellars, W., 478
semantic, 48, 133, 136–139, 141, 143, 146, 205–207, 209, 210, 213, 218, 220
  analysis, 145
  case against actual infinity, 548, 565
  completeness, 501
  consequence, 55, 821, 827
  formal, 41
  interpretation, 67
  metaparadox, 139, 141
  paradox, 143
  relations, 133, 146
  semantically inert, 1115
  semantic web, 49
  valuations, 821
semantics of relevance logics, *see* four-valued, dual gaggle, relational, operational and semi-lattice semantics
semi-lattice semantics 746–748
sense, 915, 917, 918
sentence negation, 1092
sentence schema, 55
sentence-contradictions, 1095
sentential contradictions, 1095, 1101
sentential logic, 210
sentential modal operator, 1115
separation axiom, 532, 541, 544
sequent calculus, 737, 739, *see* consecution calculus
sequents, 508, 509
set theory, 41, 50, 54, 133, 145, 431
  paradoxes of, 413
sets, 528, 530–546
Shaffer, J., 151
Shapiro, L., 405
Shapiro, S., 372, 923
Sheard, M., 405
Shehtman, V., 945
Shelah, S., 607
Ship of Thesus, 168
Shoenfield, J. R., 54, 542–543, 843
Sider, T., 979
Sieg, W., 414, 415, 418, 421, 423, 429
Sigwart, C., 60

similarity, 530, 956, 1130
Simmons, K., 386, 387
Simons, P., 1091
simple pragmatic structure, 852, 853
simple types, 723, 748–754, 960, *see* combinator
Simpson, S., 439
Sinn, 1062, 1107–1109, 1118, 1119
situation, 944, 951
  ethicists, 164
  semantics, 82
Skolem, T., 499–502, 506, 510, 511, 516, 518, 550–551, 587, 615
  Skolem's Paradox, 587, 615–647
Skvortsov, D., 945
Skyrms, B., 137, 384, 405, 980
Slater, H., 1077, 1085, 1097
slingshot argument, 913, 920
slippery slope, 168
small, 602
small matrices, 206
Smiley, T., 220
  Smiley 8-value, 216, 221, 224
  Smiley matrix, 213, 220
Smith, B. H., 568, 569, 1082
smooth infinitesimal analysis, 528
Smoryński, C., 433
Smullyan, R., 331, 339, 729, 731
Snow, C. P., 132
Soames, S., 330, 382
Sochor, A., 529, 537, 562
Socrates, 167, 169
Solovay, R. M., 435
Sorensen, R. A., 155, 159, 161
sorites, 155, 157
sound, 1092
soundness
  of modal logics, 743, 757
  of relevance logics, 731, 744, 746, 748, 759, 760, 765, 768, 775, 777, 781, 782
Souza, E. G., 831
species of structure, 849
spectrum, 588, 595
speech acts
  assertives premissing, 103–105, 106
  concluding, 106
  commissives, 103–105
  declaratives, 103–105
  directives, 103–105
  expressives, 103–105
  pragmatics, 1069
speed, 58
speed up, 435
square of opposition, 131, 133, 134
squeeze lemma, 766
St. Anselm, 146
Stalnaker, R., 1127, 1128
Stalnaker–Lewis semantics, 1131
standard semantics, 686, 923
standard translation, 947
Stanovich, K. E., 59
state-of-affairs, 950
statistics, 42
Stecker, R., 1077
Stegmüller, W., 54
Stein, H., 414
Steiner, M., 414
Stenning, K., 51
stipulation, 1071, 1073, 1089, 1103–1105, 1107, 1110, 1111, 1114
Stokes, G. G., 792
Stone, W. H., 1088
  Stone representation theorem, 1088
Stove, D., 1064
stratification of constructions, 557–558
Strawson, P., 170
strict finitism, 206, 210, 224
strict implication, 926, 952
strong AI, 1072
Strong Kleene, 333, 335–342, 344–346, 350–354, 363, 370, 375
strong negation, 804, 861, 887
structural content, 1013–1014, 1018–1021
structural rules, 79
structuralist, 484, 1015–1018

structuralist thesis, 1016–1018
structure, 589
subformula property, 739, 740, 784
subjective probability, 890
Subrahmanian, V. S., 858, 859, 874
substantiation, 1105
substitution principles, 958
  first-order, 959
  second-order, 959
substitutional, 1114
substitutional quantification, 1080, 1107, 1108
substructure, 589
S, the successor function, 547
suitable for E, 214
suitable for a system, 210, 214, 220, 225
Sundholm, G., 556
supertasks, 565, 571–575
supervaluational logic, 1077, 1097
supervaluationism, 163
Suppes, P., 849
surveyable, 206, 210
syllogism, 44–46, 58, 67, 133, 135, 828
  argument, 133
  logic, 134, 135
Sylvan, R., 724, 758, 763, 774, 778, 782
symbolic logic, 135, 136
symbolisms, 135
syntactic consequence
  proper, 857
syntax, 49, 133
syntax v. semantics, 47, 48
system E, 205, 208, 211, 213, 215, 217, 220
system imprisonment, 66
system R, 215
Słupecki, J., 682

T-schema, 1115
T^F, 1072, 1094, 1096, 1101, 1103, 1108, 1112, 1115–1117
T (ticket entailment), 731–774
T, F, 916, 918, 919, 923
t (truth), 738, 739, 740, 755, 757, 760, 776
t-norm algebra, 698, 699
tableaux techniques, 213
tacit knowledge, 1111
Taff's Well, 160
Tait, W. W., 422, 425–427, 440, 525, 549, 551, 1110
Takeuti, G., 432, 436
taking, 1089
target properties, 1062, 1066, 1067
targets, 1064, 1066
Tarski*, 382, 383, 386, 387, 398, 404
Tarski, A., 49, 53–57, 137, 138, 142, 327, 328, 331, 334, 344, 382, 456, 498, 506, 507, 519, 567, 568, 593, 594, 710, 712, 850, 853, 918, 964
  Tarski's theory of truth, 143
  Tarskian, 138
  Tarskian hierarchy, 141
  Tarskian semantic hierarchy, 143
  Tarskian structure of metalanguage, 143
Tarski-Knaster theorem, 868
tautology, 139, 144–146, 150
tautology of propositional logic, 141
Taylor, K., 1077
teaching logic, 53, 54, 57–59
temporal assymetry of causation, 1130
temporal logic of programs, 955
textbooks, 53, 54
Tharp, L. H., 544, 1104
the lattice FOUR, 866, 877, 881
the lattice SQUARE, 866
The Word, 1103
the world, 1121
theorem
  compactness, 1092
  deduction, 1092
  extrapolation, 1092
theorem proving methods, 876

theory, 68, 808
  complementary, 896
  confidence, 889
  consistent, 791, 809
  descriptions, 487
  dynamics, 85
  games, 1088
  inconsistent, 791, 792, 809
  light rays, 1018–1020
  objects, 1078
  paraconsistent, 809
  possibility, 890
  trivial, 791
  truth, 1107
  types, 488, 498, 510, 921
thin posit, 1107
things, 973
Thistlewaite, P. B., 778, 782
Thom, P., 45
Thomas, R. S. D., 1104
Thomason, S. K., 964
Thomson, J., 568, 571, 573
thread, 977
three-valued logic, 871, 1078
Tichý, P., 929, 935, 936
ticket entailment, 724, 729–774, *see* axiomatic and consecution calculi
  D-completeness of $\mathbf{T}_{\rightarrow}$, 754
  fde fragment of, 729–731
  and $\mathbf{HRM}_z$-terms, 750, 751
  Fitch-style calculus for, 746–747
  implicational fragment of, 733–737
  intensional fragment of, 737–743
  positive fragment of, 754–762
  semantics of, 743, 746, 748, 757–768
Tierney, M., 50
Tiles M., 1104
token-type distinction, 547
top-down theories of constructive proof, 556–558
topic neutrality, 162
toposes, 50
Toraldo di Francia, G., 849–851
Torretti, R., 1104
Toulmin, S. E., 122, 1081
tower, 965
trace (of an object), 973
trace function, 973
transfer theorem, 588
transfictive discourse, 1075
transfictive sentence, 1075
transfinite, 525, 526, 533–536
transfinite iteration, 539, 541
transfixive, 1117
transform of a formula, 813
transitive, 1129, 1132, 1142
  model, 620, 633–636
translation, 160
transworld accessibility relation, 147
transworld identity, 945
Treaty of Lausanne, 160
Triad rule, 212–214, 224, 225
trivial, 895, 897
trivial theory, 797
  seetheory: trivial, 791
trivialization of a calculus, 806
Troelstra, A. S., 548, 554, 559, 563
Trout, J. D., 164
trumping preemption, 1133, 1135, 1138
truth, 59, 136, 137, 139, 140, 143–146, 474, 913, 918, 919, 946, 1064, 1067–1069
  contextual, 384–387
  determinate, 349, 351, 358–360, 366, 398
  dtruth, 325–331, 333, 334, 337, 339–348, 350–354, 356–358, 360–372, 375, 377–381, 383, 385, 388, 392
  degree functions, 677
  degrees, 677
  fiction, 1120
  indexical, 386, 387
  L-truth, 964
  ordering, 696

  revision-theoretic, 392, 393, 395, 396
  situational, 389–391
  stipulation, 1104, 1116
  strong (-er), 350, 351, 353, 356–360, 362, 369, 372, 377, 381
  table, 137
  value gaps, 137, 141, 163
  value predications, 138
  values, 208, 210, 913, 915, 916, 919–921, 924
  valueless, 1076
Tupailo, S., 436
Turing, A. M., 411, 498, 514, 515
  Turing computability, 515
  Turing machine, 514, 515, 794
twin worlds, 949
two cardinal theorems, 600
two-sorted type theory, 935
two-variable fragment, 972
types, 455

ultrafilter, 951
ultraproduct, 506, 519
unary operation, 218
uncertainty, 874
undecidability, 498, 502, *see* decidability
undecidable predicate logic, 215
undecidable statements, 503, 506
undefinability, 498, 503
Unger, P., 163
universal, 479
  instantiation, 961
  machine, 515
  modality, 948
  set, 833
unsolvability, 514
updating, 90
upward Löwenheim-Skolem-Tarski theorem, 594
Urbas, I., 871
urelement, 517, 518
urelements, 517
Urmson, J. O., 1061
Urquhart, A., 215, 743, 746, 769, 770, 778, 781
use theory, 163

Väänänen, J., 607
Vago, C., 859
vague propositions, 884
vagueness, 793, 885
Vaihinger, H., 1104
valid, 60, 134, 145, 205, 208, 212, 215, 217, 679
  deductive inference, 139
  inference, 133
  matrix, 220
  semantics, 140
valuation, 211, 821, 826, 886, 946
  scheme, 334
van Dalen, D., 548, 554, 559
van Fraassen, B., 340, 385, 900
van Stigt, W. P., 546, 564
variable assignment, 211
variable sharing, 772, 781, 782
  in relevance logics, 724, 731, 772, 782
  missing in classical logic, 727
Varzi, A. C., 329, 569
Vasiliev, N. I., 792–794
  imaginary geometry, 792
Vaught, R. L., 594, 599, 601, 606
Veblen, O., 497
Venn, 133, 134
  Venn diagramming, 135
verification criterion, 459
verificationism, 986
vicious circle principle, 537, 540
Vienna Circle, 474
Vieu, L., 569
Visser, A., 339, 367, 403, 435
vocabulary, 589
von Neumann, J., 417, 418, 430, 431, 441, 449, 505, 517, 527, 794, 892, 1088
  von Neumann Hierarchy, 596

Vopěnka, P., 529, 537, 562

W (combinator), 733, 736–738, 750, 751
  simple type of, 733, 745, 769, 777
W$_*$ (combinator)—seoM, 785
Wagner, S., 1104
Wajsberg, M., 682
Walton, D. N., 117, 122–123, 126
  Walton's dilemma, 1084
  Waltonian truth, 1084
Walton, K., 1061, 1070, 1073, 1074, 1082, 1083, 1085, 1086, 1088, 1094, 1101
Wang, H., 542–544
  Wang's paradox, 157
warning rule, 884, 890–892
weak
  absolute, 609
  AI, 1072
  Kleene, 335, 340
  realism, 1074
  second-order logic, 592
Webb, J., 414
Weierstrass, K., 528
Weinberg, S., 577, 578
Weinstein, S., 564
well-behaved formulas, 799
well-foundedness, in supertasks, 574
well-individuated object, 1107, 1108
well-ordered sets, 530, 535
well-ordering number, 605
Wesson, P. S., 576
Weyl, H., 53, 56, 415, 416, 428–430, 498, 531, 537, 540, 550, 552, 565, 573
Whitehead, A. N., 50, 411, 414, 449, 498, 537
William of Ockham, 48
Williamson, T., 155
Wilson chamber, 851
Wittgenstein, L., 142–145, 167–169, 463, 565, 948, 1088
Wittgensteinian, 206
Wolf, R. G., 874
Woltersdorff, N., 1061, 1070, 1099
Woodger, J. H., 138
Woodruff, P., 330, 333, 335, 338, 340
Woods, J., 126, 1061, 1068, 1075, 1079, 1080, 1091–1093, 1095, 1097, 1103, 1112, 1117, 1121
world-line matrices, 216, 218–220
world-mirror, 981
worldbound, 976
worlds, 944, 946, 1097
Wright, C., 169, 384, 441, 465, 559
Wu, J., 859

Xia, Z., 576

Yablo, S., 351, 356, 367, 574, 1136–1139
Yamashita, M., 872
Yaqūb, A., 405
Yessenin-Volpin, A. S., 559, 561

Zach, R., 415, 418, 425, 428
Zadeh, L., 876, 890
Zalta, E. N., 1070, 1075, 1092
Zeno, 565–571, 576
Zermelo, E., 50, 413–414, 449, 498–500, 503, 516, 532, 533, 540, 542, 587, 1071
  Zermelo's axiomatization of set theory, 497
Zermelo–Fraenkel set theory (ZFC), 517, 532, 830, 833, 842, 863, 892, 1071, 1073, 1105, 1107
  hyphenation, 1107
  operator, 1107
Zimmerman, T. E., 935
Zink, J., 528, 567
Zorn's lemma, 767